**ĐẠI HỌC ĐÀ NẴNG**

**TRƯỜNG ĐẠI HỌC KINH TẾ**

🙣 🕮 🙡

****

**BÁO CÁO CHUYÊN ĐỀ**

**THỰC TẬP TỐT NGHIỆP**

**ỨNG DỤNG BUSINESS INTELLIGENCE (BI)   
PHÂN TÍCH MỨC ĐỘ HÀI LÒNG CỦA KHÁCH HÀNG VỀ SỬ DỤNG DỊCH VỤ CỦA CHUỖI RỬA XE   
THÔNG MINH VIETWASH**

SVTH: Hồ Ngọc Hoàng Vũ

Lớp: 42K21

GVHD: ThS. Cao Thị Nhâm

***Đà Nẵng, 05/2020***

NHẬN XÉT CỦA DOANH NGHIỆP THỰC TẬP

1. ***Về ý thức và tinh thần trách nhiệm của sinh viên khi thực tập tại doanh nghiệp:***

1. ***Về năng lực và khả năng thích nghi của sinh viên với môi trường doanh nghiệp:***

1. ***Kiến nghị của Doanh nghiệp đối với Nhà trường để cải thiện chất lượng đào tạo:***

LỜI CẢM ƠN

Lời đầu tiên, em xin chân thành gửi lời cảm ơn tới cô Cao Thị Nhâm – Giảng viên khoa Thống kê – Tin học trường Đại học Kinh tế Đà Nẵng đã nhiệt tình hướng dẫn để em có thể hoàn thành tốt bài báo cáo này.

Em xin chân thành cảm ơn tất cả thầy cô khoa Thống Kê – Tin Học cùng các thầy cô trường Đại Học Kinh Tế - Đại Học Đà Nẵng đã tận tình giảng dạy, truyền đạt nhiệt huyết cho em trong suốt bốn năm học qua.

Em xin gửi lời cảm ơn chân thành tới anh Lê Khả Chung – Gíam đốc công ty EFOX SOLUTION đã tạo điều kiện cho em thực tập cũng như cung cấp số liệu để em hoàn thành tốt bài báo cáo tốt nghiệp cuả mình.

Em xin chân thành cảm ơn!

**Sinh viên**

**Hồ Ngọc Hoàng Vũ**

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi. Các số liệu, kết quả nêu trong báo cáo là hoàn toàn trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

*Sinh viên thực hiện*

**Hồ Ngọc Hoàng Vũ**

MỤC LỤC

[NHẬN XÉT CỦA DOANH NGHIỆP THỰC TẬP ii](#_Toc41558765)

[LỜI CẢM ƠN iv](#_Toc41558766)

[LỜI CAM ĐOAN v](#_Toc41558767)

[MỤC LỤC. vi](#_Toc41558768)

[DANH MỤC TỪ VIẾT TẮT ix](#_Toc41558769)

[DANH MỤC BẢNG BIỂU x](#_Toc41558770)

[DANH MỤC HÌNH ẢNH xi](#_Toc41558771)

[MỞ ĐẦU . 1](#_Toc41558772)

[**1)** **Lý do chọn đề tài** 1](#_Toc41558773)

[**2)** **Mục đích nghiên cứu** 1](#_Toc41558774)

[**3)** **Phạm vi và phương pháp nghiên cứu** 2](#_Toc41558775)

[**4)** **Kết quả đạt được** 2](#_Toc41558776)

[**5)** **Cấu trúc báo cáo** 2](#_Toc41558777)

[CHƯƠNG 1: GIỚI THIỆU VỀ CÔNG TY EFOX SOLUTION VÀ CHUỖI RỬA XE THÔNG MINH VIETWASH 3](#_Toc41558778)

[1.1. Giới thiệu về công ty EFOX SOLUTION ( Đơn vị thực tập) 3](#_Toc41558779)

[1.2. Giới thiệu về chuỗi rửa xe thông minh VIETWASH 4](#_Toc41558780)

[CHƯƠNG 2: CƠ SỞ LÝ THUYẾT 6](#_Toc41558781)

[2.1. Tổng quan về Business Intelligence (BI) 6](#_Toc41558782)

[2.1.1. Business Intelligence (BI) 6](#_Toc41558783)

[2.1.2. BI có lợi ích gì? 6](#_Toc41558784)

[2.1.3. Những thành phần chủ yếu của Business Intelligence (BI) 6](#_Toc41558785)

[2.2. Tổng quan về khai phá dữ liệu (data mining) 8](#_Toc41558786)

[2.2.1. Khái niệm 8](#_Toc41558787)

[2.2.2. Các phương pháp khai phá dữ liệu 8](#_Toc41558788)

[2.2.3. Quy trình 8](#_Toc41558789)

[2.3. Tổng quan về luật kết hợp 9](#_Toc41558790)

[2.3.1. Giới thiệu về luật kết hợp 9](#_Toc41558791)

[2.3.2. Một số tính chất liên quan đến các hạng mục phổ biến: 11](#_Toc41558792)

[2.3.3. Luật kết hợp: 11](#_Toc41558793)

[2.3.4. Thuật toán Apriori 12](#_Toc41558794)

[CHƯƠNG 3: GIỚI THIỆU VỀ CÁC CÔNG CỤ 17](#_Toc41558795)

[3.1. Tổng quan về công cụ trực quan hóa Power BI 17](#_Toc41558796)

[3.1.1. Giới thiệu về Power BI 17](#_Toc41558797)

[3.1.2. Các bộ phận của Power BI 18](#_Toc41558798)

[3.1.3. Công dụng của Power BI 18](#_Toc41558799)

[3.2. Tổng quan về Microsoft SQL Server 2014 19](#_Toc41558800)

[3.2.1. Giới Thiệu SQL Server 19](#_Toc41558801)

[3.2.2. Các thành cơ bản trong SQL Server 19](#_Toc41558802)

[3.3. Microsoft SQL Server Analysis Services (SSAS) 21](#_Toc41558803)

[3.4. Môi trường triển khai ứng dụng (BIDS) 22](#_Toc41558804)

[CHƯƠNG 4: PHÂN TÍCH PHẢN HỒI VỀ VIỆC SỬ DỤNG DỊCH VỤ TẠI CHUỖI RỬA XE THÔNG MINH VIETWASH 23](#_Toc41558805)

[4.1. Thực trạng và yêu cầu 23](#_Toc41558806)

[4.2. Đề xuất qui trình 23](#_Toc41558807)

[4.2.1. Thu thu thập dữ liệu 24](#_Toc41558808)

[4.2.2. Thiết kế dữ liệu 26](#_Toc41558809)

[4.2.3. Cài đặt dữ liệu vào hệ thống 26](#_Toc41558810)

[4.2.4. Trực quan hóa dữ liệu 28](#_Toc41558811)

[4.2.5. Xây dựng mô hình khai phá dữ liệu trên BIDS 31](#_Toc41558812)

[4.2.6. Ứng dụng luật kết hợp phân tích phản hồi khách 35](#_Toc41558813)

[CHƯƠNG 5: KẾT QUẢ PHÂN TÍCH 45](#_Toc41558814)

[5.1. Kết quả báo cáo BI 45](#_Toc41558815)

[5.1.1. Thống kê khách hàng theo loại xe 45](#_Toc41558816)

[5.1.2. Thống kê khách hàng theo đánh giá 46](#_Toc41558817)

[5.1.3. Thống kê khách hàng theo mức độ quay lại 47](#_Toc41558818)

[5.1.4. Thống kê khách hàng theo giới tính và ngành nghề 48](#_Toc41558819)

[5.1.5. Thống kê khách hàng theo mức độ quay lại và ngành nghề 49](#_Toc41558820)

[5.1.6. Thống kê khách hàng theo mức độ quay lại và thu nhập trung bình 50](#_Toc41558821)

[5.1.7. Thống kê khách hàng theo đánh giá và nhóm tuổi 52](#_Toc41558822)

[5.1.8. Thống kê khách hàng theo mức độ quay lại và đánh giá 53](#_Toc41558823)

[5.2. Kết quả báo cáo luật kết hợp 54](#_Toc41558824)

[5.2.1. Kết quả mô hình AR1 54](#_Toc41558825)

[5.2.2. Kết quả mô hình AR2 56](#_Toc41558826)

[5.2.3. Kết quả mô hình AR3 57](#_Toc41558827)

[5.2.4. Kết quả mô hình AR4 58](#_Toc41558828)

[5.2.5. Kết quả mô hình AR5 59](#_Toc41558829)

[5.2.6. Itemsets 59](#_Toc41558830)

[5.2.7. Mạng phụ thuộc (Dependency Network) 60](#_Toc41558831)

[5.3. Tổng kết phân tích 62](#_Toc41558832)

[KẾT LUẬN. 63](#_Toc41558833)

[TÀI LIỆU THAM KHẢO lxiv](#_Toc41558834)

[PHỤ LỤC. lxv](#_Toc41558835)

DANH MỤC TỪ VIẾT TẮT

|  |  |
| --- | --- |
| **Từ viết tắt** | **Diễn giải** |
| SQL | Structure Query Language |
| BI | Business Intelligence |
| SSAS | SQL Server Analysis Services |
| BIDS | Business Intelligence Development Studio |

DANH MỤC BẢNG BIỂU

[Bảng 1 Kiểu dữ liệu và kích thước 26](#_Toc41534719)

DANH MỤC HÌNH ẢNH

[Hình 1.1 Chuỗi rửa xe thông minh VIETWASH 4](#_Toc41534738)

[Hình 2.1 Những thành phần chủ yếu của Business Intelligence (BI) 6](#_Toc41534739)

[Hình 2.2 Quy trình khai phá dữ liệu 8](#_Toc41534740)

[Hình 3.1 Môi trường làm việc của Power BI 17](#_Toc41534741)

[Hình 3.2 Các thành phần cơ bản trong SQL 19](#_Toc41534742)

[Hình 4.1 Đề xuất quy trinhg khai phá dữ liệu 23](#_Toc41534743)

[Hình 4.2 Thu thập dữ liệu 24](#_Toc41534744)

[Hình 4.3 Tạo Data base mới trong Microsoft SQL 26](#_Toc41534745)

[Hình 4.4 Import dữ liệu vào Database 27](#_Toc41534746)

[Hình 4.5 Import dữ liệu thành công 27](#_Toc41534747)

[Hình 4.6 Dữ liệu trong Database 28](#_Toc41534748)

[Hình 4.7 Sử dụng công cụ Power BI 28](#_Toc41534749)

[Hình 4.8 Tải dữ liệu vào Power BI 29](#_Toc41534750)

[Hình 4.9 Trình điều hướng sau khi tải dữ liệu 29](#_Toc41534751)

[Hình 4.10 Màn hình làm việc sau khi tải dữ liệu 30](#_Toc41534752)

[Hình 4.11 Tạo project mới 31](#_Toc41534753)

[Hình 4.12 Tạo Data Source Wizard 31](#_Toc41534754)

[Hình 4.13 Tạo Data Source Wizard 32](#_Toc41534755)

[Hình 4.14 Tạo Data Source Wizard 32](#_Toc41534756)

[Hình 4.15 Tạo Data Source Wizard 33](#_Toc41534757)

[Hình 4.16 Tạo Data Source View 33](#_Toc41534758)

[Hình 4.17 Tạo Data Source View 34](#_Toc41534759)

[Hình 4.18 Tạo Data Source View 34](#_Toc41534760)

[Hình 4.19 Tạo Data Source View thành công 35](#_Toc41534761)

[Hình 4.20 Các bước thực thi luật kết hợp 35](#_Toc41534762)

[Hình 4.21 Giao diện tạo Mining Structure 36](#_Toc41534763)

[Hình 4.22 Tạo Mining Structure 36](#_Toc41534764)

[Hình 4.23 Lựa chọn mô hình luật kết hợp 37](#_Toc41534765)

[Hình 4.24 Chọn thuộc tính để training (Mô hình AR1) 37](#_Toc41534766)

[Hình 4.25 Đặt tên mô hình luật kết hợp (Mô hình AR1) 38](#_Toc41534767)

[Hình 4.26 Chọn thuộc tính để Training (Mô hình AR2) 38](#_Toc41534768)

[Hình 4.27 Đặt tên mô hình luật kết hợp (Mô hình AR2) 39](#_Toc41534769)

[Hình 4.28 Chọn thuộc tính training (Mô hình AR3) 39](#_Toc41534770)

[Hình 4.29 Đặt tên mô hình luật kết hợp (Mô hình AR3) 40](#_Toc41534771)

[Hình 4.30 Chọn thuộc tính để training (Mô hình AR4) 40](#_Toc41534772)

[Hình 4.31 Đặt tên mô hình luật kết hợp (Mô hình AR4) 41](#_Toc41534773)

[Hình 4.32 Chọn thuộc tính để training (Mô hình AR5) 41](#_Toc41534774)

[Hình 4.33 Đặt tên mô hình luật kết hợp (Mô hình AR5) 42](#_Toc41534775)

[Hình 4.34 Thiết lập tham số mô hình luật kết hợp 44](#_Toc41534776)

[Hình 5.1 Thống kê khách hàng theo loại xe 45](#_Toc41534777)

[Hình 5.2 Thống kê khách hàng theo đánh giá 46](#_Toc41534778)

[Hình 5.3 Thống kê khách hàng theo mức độ quay lại 47](#_Toc41534779)

[Hình 5.4 Thống kê khách hàng theo giới tính và ngành nghề 48](#_Toc41534780)

[Hình 5.5 Thống kê khách hàng theo mức độ quay lại và ngành nghề 49](#_Toc41534781)

[Hình 5.6 Thống kê khách hàng theo mức độ quay lại và thu nhập trung bình 50](#_Toc41534782)

[Hình 5.7 Thống kê khách hàng theo đánh giá và nhóm tuổi 52](#_Toc41534783)

[Hình 5.8 Thống kê khách hàng theo mức độ quay lại và đánh giá 53](#_Toc41534784)

[Hình 5.9 Kết quả mô hình AR1 54](#_Toc41534785)

[Hình 5.10 Kết quả mô hình AR2 56](#_Toc41534786)

[Hình 5.11 Kết quả mô hình AR3 57](#_Toc41534787)

[Hình 5.12 Kết quả mô hình AR4 58](#_Toc41534788)

[Hình 5.13 Kết quả mô hình AR5 59](#_Toc41534789)

[Hình 5.14 Itemsets của mô hình 60](#_Toc41534790)

[Hình 5.15 Mạng phụ thuộc (Mô hình AR1) 60](#_Toc41534791)

[Hình 5.16 Mạng phụ thuộc (Mô hình AR2) 61](#_Toc41534792)

[Hình 5.17 Mạng phụ thuộc (Mô hình AR3) 61](#_Toc41534793)

[Hình 5.18 Mạng phụ thuộc (Mô hình AR4) 61](#_Toc41534794)

[Hình 5.19 Mạng phụ thuộc (Mô hình AR5) 62](#_Toc41534795)

MỞ ĐẦU

1. **Lý do chọn đề tài**

Ở một quốc gia mà số lượng phương tiện cá nhân trên đầu người khá cao như Việt Nam thì nhu cầu rửa xe của người dân tất yếu sẽ rất lớn. Tuy nhiên, hiện nay đa số dịch vụ rửa xe là nhỏ lẻ và thô sơ, chưa có đầu tư về mặt công nghệ. Do vậy VIETWASH một trong những doanh nghiệp đi đầu trong việc hình thành chuỗi rửa xe thông minh đầu tiên tại Việt Nam nhằm cải tiến quy trình rửa xe một cách nhanh chóng và hoàn toàn tự động.

Vì rửa xe thông minh là một dịch vụ mới chưa được phổ biến, đa số chủ phương tiện đều đã quen sử dụng dịch vụ ở các tiệm rửa xe nhỏ lẻ. Do vậy cần khảo sát ý kiến khách hàng ban đầu để đánh giá ,phân tích, từ đó đưa ra những quyết đinh, kế hoạch phù hợp nhằm nâng cao chất lượng dịch vụ cũng như cũng đánh giá của khách hàng trong tương lai, ngoài ra cải thiện chính sách chăm sóc khách hàng , tăng lợi thế cạnh canh cho doanh nghiệp.

Từ kiến thức của bản thân và dữ liệu do công ty EFOX Solution cung cấp, tôi mong muốn tiến hành những đánh giá về mức độ hài lòng của khách hàng khi sử dụng dịch vụ tại chuỗi rửa xe VIETWASH, nhờ vào các công cụ khai phá dữ liệu để khai thác những thông tin hữu ích từ kho dữ liệu của công ty, qua đó có thể đưa ra những định hướng đúng đắn trong quá trình kinh doanh cũng như nâng cao lòng tin của khách hàng khi sử dụng dịch vụ.

Vì vậy tôi đã chọn đề tài “Ứng dụng business intelligence (BI) phân tích mức độ hài lòng của khách hàng về sử dụng dịch vụ của chuỗi rửa xe thông minh VIETWASH” làm đề tài báo cáo tốt nghiệp cho mình.

1. **Mục đích nghiên cứu**

Bài báo cáo này được thực hiện nhằm tìm hiểu về Business Intelligence, ứng dụng luật kết hợp trong khai phá dữ liệu để ứng dụng vào viêc phân tích mức độ hài lòng của khách hàng về việc sử dụng dịch vụ tại chuỗi rửa xe thông minh VIETWASH. Từ đó đưa ra các định hướng, kế hoạch hợp lý cho doanh nghiệp.

1. **Phạm vi và phương pháp nghiên cứu**

Nghiên cứu về phản hồi của khách hàng khi sử dụng dịch vụ tại các điểm của chuổi rửa xe VIETWASH trên địa bàn TP HCM ( khoảng 50 điểm).

Phương pháp phân tích thông tin bằng các dữ liệu đã được thu thập tại chuỗi rửa xe thông minh VIETWASH.

1. **Kết quả đạt được**

Phân tích được mức độ hài lòng của khách hàng khi sử dụng dịch vụ và tìm ra được các luật kết hợp có liên quan để đánh giá được mức độ hài lòng khách hàng.

1. **Cấu trúc báo cáo**

**CHƯƠNG 1**: Giới thiệu về công ty EFOX SOLUTION và chuỗi rửa xe thông minh VIETWASH.

**CHƯƠNG 2**: Cơ sở lý thuyết.

**CHƯƠNG 3:** Giới thiệu về các công cụ

**CHƯƠNG 4:** Phân tích phản hồi của khách hàng về việc sử dụng dịch vụ tại chuỗi rửa xe thông minh VIETWASH.

**CHƯƠNG 5:** Kết quả phân tích.

# GIỚI THIỆU VỀ CÔNG TY EFOX SOLUTION VÀ CHUỖI RỬA XE THÔNG MINH VIETWASH

## Giới thiệu về công ty EFOX SOLUTION ( Đơn vị thực tập)

EFOX Solutions là một công ty chuyên về công nghệ thông tin và tư vấn giải pháp công nghệ, marketing. EFOX luôn cam kết cung cấp dịch vụ trọn gói cho các khách hàng trong nước và quốc tế, nhờ đó, có thể giúp khách hàng giảm thiểu các chi phí hành chính để tập trung vào những hoạt động kinh doanh then chốt của khách hàng.

Đội ngũ nhân sự chuyên nghiệp của EFOX Solutions gồm gần 30 lập trình viên, chuyên gia marketing, designer và nhân viên phối hợp làm việc với nhau tại ba văn phòng ở thành phố Hồ Chí Mính, Đà Nẵng và Vương Quốc Anh để phục vụ tất cả các vấn đề về công nghệ và marketing cho các khách hàng trong và ngoài nước.

Với thế mạnh là khả năng giải quyết nhanh các vấn đề về công nghệ cũng như là marketing trong và ngoài nước cho các doanh nghiệp hay cá nhân, EFOX chính là một sự lựa chọn hoàn hảo đối với các công ty, các tập đoàn và những tổ chức khá có quy mô vừa và lớn.

**Giai đoạn thành lập và mở rộng - 2015**

Giám đốc Lê Khả Chung đã thành lập nên EFOX Solutions vào tháng 08/2015 tại Thành Phố Hồ Chí Minh và công ty chỉ tập trung vào các lĩnh vực thiết kế website và outsourcing cho các doanh nghiệp nước ngoài. Trong 2 năm kế tiếp, EFOX đã tăng trưởng đáng kể và được nhiều khách hàng quốc tế đánh giá cao.

**Giai đoạn tăng trưởng - 2017**

Vào thời gian năm 2017, EFOX mở rộng thêm lĩnh vực thiết kế và phát triển ứng dụng di động. Số lập trình viên và nhân viên hành chính lúc này lên tới 20 người. Cũng trong thời gian này EFOX Solutions mở thêm văn phòng tại Vương Quốc Anh, nhằm mở rộng thêm thị trường và hỗ trợ tốt hơn cho khách hàng khu vực Châu Âu. Vào thời gian này, EFOX Solutions đã tiếp nhận 1 số khách hàng là những doanh nghiệp lớn trong nước.

**Giai đoạn củng cố và đột phá 2019 – 2020**

Với sự phát triển nhanh chóng do nhận được nhiều sự tin tưởng từ những cá nhân, doanh nghiệp lớn trong nước, EFOX đã mở thêm văn phòng tại khu vực Đà Nẵng nhằm hỗ trợ và phát triển các khách hàng tốt hơn, ngoài ra với xứ mệnh xây dựng đội ngũ nhân viên xuyên suốt 3 miền của EFOX Solutions tạo nên một gia đình với đa bản sắc từ trong và ngoài nước. Lúc này EFOX Solutions đã mở rộng thêm các lĩnh vực khác như: Marketing chuyên sâu, đào tạo marketing doanh nghiệp, xây dựng và phát triển thương hiệu cá nhân, doanh nghiệp, SEO top google…

CÁC DỊCH VỤ :

* Thiết kế ứng dụng di động
* Lập trình chức năng ứng dụng di động
* Thiết kế giao diện website thương mại điện tử
* Tùy biến website thương mại điện tử sẵn có
* Phát triển hệ thống thanh toán, vận chuyển của website thương mại điện tử
* Dịch vụ SEO

## Giới thiệu về chuỗi rửa xe thông minh VIETWASH

Cửa hàng rửa xe tự động VIETWASH đầu tiên trong chuỗi rửa xe thông minh VIETWASH được đưa vào hoạt động 11/2014, với sự hợp tác của xí nghiệp bán lẻ xăng dầu Petrolimex Sài Gòn. Hiện nay, số lượng điểm rửa xe chuỗi rửa xe tự động VIETWASH đã tăng lên thành hơn 50 điểm, “phủ sóng” rộng khắp tại TP. Hồ Chí Minh, Bình Dương, Đồng Nai, Đà Nẵng,...Sự thành công của mô hình này được tạo nên từ chính những yếu tố khác biệt.



Hình . Chuỗi rửa xe thông minh VIETWASH

Khi đến với cửa hàng rửa xe thông minh VIETWASH, xe của khách hàng sẽ được giao cho nhân viên, hệ thống tiếp nhận gói dịch vụ trong 3-5 giây, in phiếu thu ngay lập tức và xe được rửa hoàn toàn tự động bằng máy, chỉ trong thời gian chưa đến 5 phút, nhanh hơn phương thức rửa xe thông thường rất nhiều. Đây là hình thức mà nhân viên tại cửa hàng chỉ cần dắt xe vào hệ thống máy rửa xe máy tự động đã được lắp đặt sẵn, vận hành chương trình rửa xe đã được cài đặt sẵn, không cần phải thao tác làm sạch bằng tay theo phương thức truyền thống.

Khi khách hàng sử dụng dịch vụ tại VIETWASH đồng nghĩa với việc hệ thống của VIETWASH sẽ cập nhật thông tin, theo dõi, tư vấn quá trình chăm sóc xe. Quy trình làm việc của chuỗi rửa xe thông minh VIETWASH rất chuyên nghiệp cho tất cả các khách hàng khi trải nghiệm dịch vụ đều cảm thấy hài lòng. Bên cạnh đó, tại các điểm rửa xe tự động VIETWASH đều kèm theo các dịch vụ như wifi miễn phí, phòng chờ hòa mát lạnh, máy massage toàn thân, massage chân, nước uống miễn phí,...mang tới sự thoải mái, hài lòng cho khách hàng. VIETWASH phục vụ khách hàng 24/7, bất kể ngày hay đêm do đó khách hàng có thể tìm tới trải nghiệm của chuỗi rửa xe thông minh VIETWASH bất cứ lúc nào.

Tất cả những yếu tố trên đã tạo nên sự khác biệt cho chuỗi rửa xe máy thông minh VIETWASH và là lợi thế để VIETWASH tự tin chinh phục thị trường, cạnh tranh với các hình thức rửa xe khác. Lợi ích to lớn nhất mà dịch vụ này mang lại cho khách hàng là thời gian và tiện ích, đó là lợi ích mà dịch vụ rửa xe truyền thống không thể có.

Với số lượng xe máy hiện có tại nước ta và chất lượng dịch vụ, sự mới mẻ của hình thức rửa xe này, mô hình chưa có đối thủ nào nhập cuộc thì hoàn toàn có thể hy vọng vào sự phát triển và gia tăng doanh thu từ chuỗi rửa xe thông minh VIETWASH trong thời gian sắp tới.

# CƠ SỞ LÝ THUYẾT

## Tổng quan về Business Intelligence (BI)

### Business Intelligence (BI)

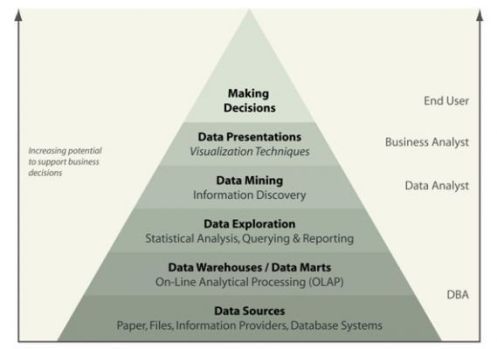
BI là một qui trình có tích hợp công nghệ mà các doanh nghiệp dùng để kiểm soát khối lượng dữ liệu khổng lồ đến từ nhiều nguồn khác nhau và khai thác nguồn dữ liệu đó giúp cho họ có thể đưa các các quyết định hiệu quả hơn trong hoạt động kinh doanh của mình. BI có mặt ở khắp các doanh nghiệp như hệ thống siêu thị, ngân hàng, viễn thông,… đó đều là những nơi cần thu thập, xử lý khối lượng dữ liệu cực lớn. Do đó BI có tính ứng dụng rất cao khi nguồn dữ liệu của doanh nghiệp sẽ lớn dần theo thời gian hoạt động. Hiện nay BI vẫn còn là một thuật ngữ khá mới ở Việt Nam, nhưng trên thế giới BI đã được ứng dụng rất nhiều vào doanh nghiệp. [1]

### BI có lợi ích gì?

BI làm tăng khả năng kiểm soát thông tin của doanh nghiệp một cách chính xác, hiệu quả từ đó có thể phân tích, khai phá tri thức giúp doanh nghiệp có thể dự đoán về xu hướng của giá cả dịch vụ, hành vi khách hàng, phát hiện khách hàng tiềm năng để đề ra các chiến lược kinh doanh phù hợp nhằm tăng khả năng cạnh tranh doanh nghiệp.

BI giúp cho các doanh nghiệp sử dụng thông tin một cách hiệu quả, chính xác để thích ứng với môi trường thay đổi liên tục và cạnh tranh khốc liệt trong kinh doanh.

### Những thành phần chủ yếu của Business Intelligence (BI)



Hình . Những thành phần chủ yếu của Business Intelligence (BI)

**Data Sources:**

Là cơ sở dữ liệu thô (thường là cơ sở dữ liệu quan hệ) đến từ nhiều nguồn khác nhau như các ứng dụng business như Human Resource Management (HRM), Customer relationship management (CRM), phần mềm bán hàng, website thương mại điện tử… Có thể là bất cứ hệ quản trị cơ sở dữ liệu nào như MySQL, Oracle, MSSQL, DB2, … Thường được thiết kế theo mô hình cơ sở dữ liệu quan hệ ( vì dạng mô hình này đang rất phổ biến trong thực tế ).

**Data Warehouse:**

Là cơ sở dữ liệu được thiết kế theo mô hình khác với CSDL quan hệ và là nơi lưu trữ dữ liệu lâu dài của tổ chức. Dữ liệu của DW chỉ có thể đọc, ko ghi hay update được và chỉ được update bởi gói ETL chuyển đổi dữ liệu từ Data Sources vào Data Warehouse.

**Data Exploration:**

**Integrating Server:**

Chịu trách nhiệm trung gian vận hành gói ETL để chuyển đổi dữ liệu từ Data Sources vào Data Warehouse.

**Analysis Server:**

Chịu trách nhiệm thực thi các Cube được thiết kế dựa trên các Dimension dữ liệu và tri thức nghiệp vụ. Cube chịu trách nhiệm nhận input data từ DW và thực thi theo nghiệp vụ định nghĩa sẵn để trả về output.

**Reporting Server:**

Thực thi các report với output nhận được từ Analysis Server.

**Data Mining:**

Là quá trình trích xuất thông tin dữ liệu đã qua xử lý (phù hợp với yêu cầu riêng của doanh nghiệp) từ Data Warehouse rồi kết hợp với các thuật toán để đưa ra ( hoặc dự đoán ) các quyết định có lợi cho việc kinh doanh của doanh nghiệp.Đây là một quá trình quan trọng trong BI, thông thường một doanh nghiệp muốn sử dụng giái pháp BI thường kèm theo về Data Mining.

**Data Presentation:**

Tạo ra các báo cáo, biểu đồ từ quá trình data mining để phục vụ cho nhu cầu của người dùng cuối. [1]

## Tổng quan về khai phá dữ liệu (data mining)

### Khái niệm

Khai phá dữ liệu là một tập hợp các kỹ thuật được sử dụng để tự động khai thác và tìm ra các mối quan hệ lẫn nhau của dữ liệu trong một tập hợp dữ liệu khổng lồ và phức tạp, đồng thời cũng tìm ra các mẫu tiềm ẩn trong tập dữ liệu đó. [2]

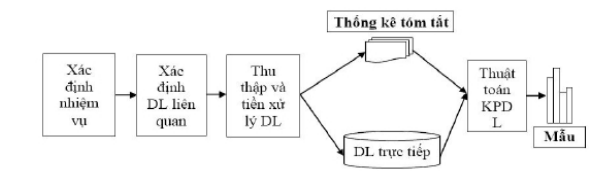
### Các phương pháp khai phá dữ liệu

* Luật kết hợp (association rules).
* Phân lớp (Classfication).
* Hồi qui (Regression).
* Trực quan hóa (Visualiztion).
* Phân cụm (Clustering).
* Tổng hợp (Summarization).
* Mô hình ràng buộc (Dependency modeling).
* Biểu diễn mô hình (Model Evaluation).
* Phân tích sự phát triển và độ lệch (Evolution and deviation analyst).
* Phương pháp tìm kiếm (Search Method).

Có nhiều phương pháp khai phá dữ liệu được nghiên cứu ở trên, trong đó có ba phương pháp được các nhà nghiên cứu sử dụng nhiều nhất đó là : Luật kết hợp, Phân lớp dữ liệu và Phân cụm dữ liệu.

### Quy trình

Quy trình khai phá dữ liệu bao gồm các bước:



Hình . Quy trình khai phá dữ liệu

* **Xác định nhiệm vụ:** Xác định chính xác các vấn đề cần giải quyết.
* **Xác định các dữ liệu liên quan:** Dùng để xây dựng giải pháp.
* **Thu nhập và tiền xử lý dữ liệu:** Thu thập các dữ liệu liên quan và tiền xử lý chúng sao cho thuật toán KPDL có thể hiểu được. Đây là một quá 16 trình rất khó khăn, có thể gặp rất nhiều các vướng mắc như: Dữ liệu phải được sao cho nhiều bản (nếu được chiết xuất vào các tệp), quản lý tập các dữ liệu, phải lặp đi lặp lại nhiều lần toàn bộ quá trình (nếu mô hình dữ liệu thay đổi), …
* **Thuật toán khai phá dữ liệu:** Lựa chọn thuật toán KPDL và thực hiện việc KPDL để tìm được các mẫu có ý nghĩa, các mẫu này được biểu diễn dưới dạng luật kết hợp, cây quyết định, … tương ứng với ý nghĩa của nó.

## Tổng quan về luật kết hợp

### Giới thiệu về luật kết hợp

Cho một tập I = {I1, I2, ..., Im} các tập m mục, một giao dịch T được định nghĩa như một tập con của các khoản mục trong I (T⊆I).

Tương tự như khái niệm tập hợp, các giao dịch không được trùng lặp, nhưng có thể nới rộng tính chất này của tập hợp và trong các thuật toán sau này, người ta đều giả thiết rằng các khoản mục trong một giao dịch và trong tất cả các tập mục khác, có thể coi chúng đã được sắp xếp theo thứ tự từ điển của các mục.

Gọi D là CSDL của n giao dịch và mỗi giao dịch được đánh nhãn với một định danh duy nhất. Nói rằng, một giao dịch T ∈ D hỗ trợ một tập X ⊆ I nếu nó chứa tất cả các item của X.

Điều này nghĩa là X ⊆ T, trong một số trường hợp người ta dùng ký hiệu T(X) để chỉ tập các giao dịch hỗ trợ cho X. Kí hiệu support(X) (hoặc sup(X), s(X)) là tỷ lệ phần trăm của các giao dịch hỗ trợ X trên tổng các giao dịch trong D, nghĩa là:



Độ hỗ trợ tối thiểu minsup là một giá trị cho trước bởi người sử dụng. Nếu tập mục X có sup(X) ≥ minsup thì ta nói X là một tập các mục phổ biến. Một tập phổ biến được sử dụng như một tập đáng quan tâm trong các thuật toán, ngược lại, những tập không phải tập phổ biến là những tập không đáng quan tâm. Các phần sau sẽ sử dụng những cụm từ khác như “X có độ hỗ trợ tối thiểu”, hay “X không có độ hỗ trợ tối thiểu” cũng để nói lên rằng X thỏa mãn hay không thỏa mãn support(X) ≥ minsup.

→Một khoản mục X được gọi là k-itemset nếu lực lượng của X bằng k, tức là |X|=k.

Một luật kết hợp có dạng R: X => Y, trong đó X, Y là tập các mục, X, Y ⊆ I và X ∩Y = ∅. X được gọi là tiên đề và Y được gọi là hệ quả của luật.

Luật X => Y tồn tại một độ tin cậy c . Độ tin cậy c được định nghĩa là khả năng giao dịch T hỗ trợ X thì cũng hỗ trợ Y. Ta có công thức tính độ tin cậy c như sau:



Tuy nhiên, không phải bất cứ luật kết hợp nào có mặt trong tập các luật có thể được sinh ra cũng đều có ý nghĩa trên thực tế. Mà các luật đều phải thoả mãn một ngưỡng hỗ trợ và tin cậy cụ thể. Thực vậy, cho một tập các giao dịch D, bài toán phát hiện luật kết hợp là sinh ra tất cả các luật kết hợp mà có độ tin cậy conf lớn hơn độ tin cậy tối thiểu minconf và độ hỗ trợ sup lớn hơn độ hỗ trợ tối thiểu minsup tương ứng do người dùng xác định. Khai phá luật kết hợp được phân thành hai bài toán con:

Bài toán 1: Tìm tất cả các tập mục mà có độ hỗ trợ lớn hơn độ hỗ trợ tối thiểu do người dùng xác định. Các tập mục thoả mãn độ hỗ trợ tối thiểu được gọi là các tập mục phổ biến.

Bài toán 2: Dùng các tập mục phổ biến để sinh ra các luật mong muốn. Ý tưởng chung là nếu gọi ABCD và AB là các tập mục phổ biến, thì chúng ta có thể xác định luật nếu AB => CD giữ lại với tỷ lệ độ tin cậy:



Nếu conf ≥ minconf thì luật được giữ lại (luật này sẽ thoả mãn độ hỗ trợ tối thiểu vì ABCD là phổ biến).

### Một số tính chất liên quan đến các hạng mục phổ biến:

***Tính chất 1***(Độ hỗ trợ của tập con):

* Với A và B là tập các mục, nếu A ⊆ B thì sup(A) ≥ sup(B)
* Điều này là rõ ràng vì tất cả các giao tác của D hỗ trợ B thì cũng hỗ trợ A.

***Tính chất 2:***

* Một tập chứa một tập không phổ biến thì cũng là tập không phổ biến.
* Nếu một mục trong B không có độ hỗ trợ tối thiểu trên D nghĩa là sup(B)< minsup thì một tập con A của B sẽ không phải là một tập phổ biến vì support(B) ≤ support(A) < minsup (theo tính chất 1)

***Tính chất 3:*** Các tập con của tập phổ biến cũng là tập phổ biến

Nếu mục B là mục phổ biến trên D, nghĩa là support(B) ≥ minsup thì mọi tập con A của B là tập phổ biến trên D vì support(A) ≥ support(B) > minsup.

### Luật kết hợp:

***Tính chất 1:***( Không hợp các luật kết hợp)

Nếu có X→Z và Y→Z trong D thì không nhất thiết X∪Y→Z là đúng

Xét trường hợp X ∩Z =∅ và các tác vụ trong D hỗ trợ Z nếu và chỉ nếu chúng hỗ trợ mỗi X hoặc Y, khi đó luật X∪Y→Z có độ hỗ trợ 0%.

Tương tự : X→Y ∧ X→Z ⇒ X→Y∪Z

***Tính chất 2:***(Không tách luật)

Nếu X∪Y→Z thì X→Z và Y→Z chưa chắc xảy ra

Ví dụ trường hợp Z có mặt trong một giao tác chỉ khi cả hai X và Y cũng có mặt, tức là sup(X∪Y)= sup(Z), nếu độ hỗ trợ của X và Y đủ lớn hơn sup(X∪Y), tức là sup(X) > sup(X∪Y) và sup(Y) > sup(X∪Y) thì hai luật riêng biệt sẽ không đủ độ tin cậy.

Tuy nhiên đảo lại: X→Y∪Z ⇒ X→Y ∧ X→Z

***Tính chất 3:*** (Các luật kết hợp không có tính bắc cầu)

Nếu X→Y và Y→Z, chúng ta không thể suy ra X→Z.

Ví dụ: giả sử T(X) ⊂ T(Y) ⊂ T(Z), ở đó T(X), T(Y), T(Z) tương ứng là các giao dịch chứa X,Y,Z, và độ tin cậy cực tiểu minconf

conf(X→Y) =conf(Y→Z)=minconf thế thì: conf(X→Y) =minconf2 < minconf vì minconf < 1, do đó luật X→Z không đủ độ tin cậy.

***Tính chất 4:***

Nếu A→(L - A) không thoả mãn độ tin cậy cực tiểu thì luật

B →(L -B) cũng không thoả mãn, với các tập mục L,A,B và B ⊆ A ⊂ L

Vì supp(B) ≥ sup(A) (theo tính chất 1) và định nghĩa độ tin cậy, chúng ta nhận được:

Cũng như vậy: Nếu có (L-C)→ C thì ta cũng có luật (L – D)→D, với D⊆C và D≠∅.

Bởi vì D⊆C nên (L - D) ⊇ (L - C), do đó sup(L - D) ≤ sup(L-C)

⇒≥ minconf

### Thuật toán Apriori

Ý tưởng thuật toán Apriori :

Apriori là một thuật giải được do Rakesh Agrawal, Tomasz Imielinski, Arun Swami đề xuất lần đầu vào năm 1993. Thuật toán tìm giao dịch t có độ hỗ trợ và độ tin cậy thoả mãn lớn hơn một giá trị ngưỡng nào đó.

Thuật toán được tỉa bớt những tập ứng cử viên có tập con không phổ biến trước khi tính độ hỗ trợ.

Thuật toán Apriori tính tất cả các tập ứng cử của tập k trong một lần duyệt CSDL. Apriori dựa vào cấu trúc cây băm. Tìm kiếm đi xuống trên cấu trúc cây mỗi khi ta chạm lá, ta tìm được một tập ứng cử viên có tiền tố chung được bao gồm trong giao dịch. Sau đó các tập ứng cử này được tìm trong giao dịch đã được ánh xạ trước đó. Trong trường hợp tìm thấy biến đếm được tăng lên 1.

Ký hiệu: Giả sử các mục trong mỗi giao dịch được lưu giữ theo trật tự từ điển. Gọi số các mục trong một tập mục là kích thước của nó và gọi tập mục có kích thước k là tập k-mục (tập k mục). Các mục trong mỗi tập mục cũng được giữ ở trật tự từ điển. Ta sử dụng các ký hiệu sau:

Lk: Tập các tập k-mục phổ biến (với độ hỗ trợ cực tiểu minsup nào đó)

Ck : Tập các tập k-mục ứng cử (các tập mục phổ biến tiềm năng)

Thuật toán Apriori

Input: CSDL D, minsup.

Output: Tập các tập mục phổ biến.

1. L1 = {Các 1 - itemset phổ biến};

2. k=2;

3. While( Lk-1! =∅ )

4. { Ck = apriori\_gen(Lk-1, minsup);// các ứng cử mới theo chương trình con ở dưới đây.

5. for( ∀ giao dịch t∈ D)

6. { Ct=Subset (Ck,t);// ứng cử viên được chứa trong t

7. for (∀ ứng cử c ∈ Ct)

8. c.count ++;

10. }

11. Lk={ c ∈ Ck⎟ c.count ≥ minsup}

12. k++;

13. }

14. Return L= ∪kLk' ;

// sinh ứng cử viên mới (\*\*)

Void apriori\_gen(Lk-1, minsup )

1. { for (∀ itemset l1∈ Lk-1)

2. for (∀ itemset l2∈ Lk-1)

3. if((L1(1)== L2(1)&&L1(2) == L2(2)&&...&& L1(k-2) == L2(k-2)) &&L1(k-1) == L2(k-1))

4. { c= L1 kết nối L2;

5. if( has\_­inrequent\_subset(c, Lk-1)) delete c;

6. else add c to Ck;

7. }

8. return Ck

9.}

Boolean has\_infrequent\_subset(c,Lk-1)

1.{ for (∀ (k-1)-subset s∈ c)

2. if(s ∉ Lk-1) return TRUE;

3. else return FALSE ;

4.}

Giải thích: Lần duyệt đầu tiên, sẽ tính số lần xuất hiện của mỗi mục để xác định các 1- itemset phổ biến. Lần duyệt thứ k (k ≥ 2) sẽ bao gồm 2 giai đoạn:

Tập phổ biến Lk-1 đã tìm thấy ở lần duyệt thứ k-1 được sử dụng để sinh ra các tập ứng cử viên Ck bằng việc sử dụng hàm Apriori\_gen.

Dựa vào CSDL, tính độ hỗ trợ của các ứng của viên trong Ck. Các ứng cử viên trong Ck mà được chứa trong giao dịch t có thể được xác định một cách hiệu quả bằng việc sử dụng cây băm được mô tả như sau:

Trong giai đoạn 2 (giai đoạn sửa, tỉa): xoá bỏ các tập c ∈ Ck sao cho một vài (k-1) – tập con của c không nằm trong Lk-1. Thủ tục này là đầy đủ bởi đối với bất kì tập nào Lk với độ hỗ trợ tối thiểu thì các tập con kích cỡ (k-1) cũng có độ hỗ trợ tối thiểu, do đó nếu ta mở rộng mỗi tập trong Lk-1 với tất cả các tập mục có thể và sau đó xoá tất cả các tập mà (k-1) – tập con của nó không nằm trong Lk-1, ta sẽ nhận được tập các tập trong Lk.

Việc kết nối là tương đương với việc mở rộng Lk-1 với mỗi mục nằm trong CSDL và sau đó xoá bỏ các tập này mà đối với nó (k-1) –itemset nhận được bằng việc xoá đi mục thứ (k-1) không nằm trong Lk-1. Ở giai đoạn này Ck ⊇ Lk . Với lập luận như vậy, giai đoạn tỉa là giai đoạn người ta xoá khỏi Ck tất cả các tập mà các (k-1) tập con của nó không nằm trong Lk-1 , cũng không xoá bất kỳ một tập nào có thể nằm trong Lk.

**Hàm Subset:** Các tập ứng cử viên Ck được lưu trữ trong một cây băm. Một nút của cây này hoặc là chứa một danh sách của các tập (nút lá) hoặc bảng băm ( một nút trong). Trong mỗi một nút trong, mỗi cụm (bucket) của bảng băm chỉ đến một nút khác. Gốc của cây băm được xem ở độ sâu là 1. Một nút trong ở độ sâu d sẽ dẫn đến nút ở độ sâu d+1. Các tập được lưu trữ trong các lá. Khi ta bổ sung thêm một tập c, ta bắt từ nút gốc và đi xuống cây cho đến khi ta chạm vào một lá. Tại một nút ở độ sâu d, ta quyết định sẽ đi theo cành nào bằng việc áp dụng hàm băm đối với mục thứ d của tập đó và theo con trỏ trong Bucket tương ứng. Tất cả các nút ban đầu được tạo ra như là nút lá. Khi số các tập trong một nút lá vượt quá ngưỡng được chọn, nút lá này được chuyển thành một nút trong.

Bắt đầu từ nút gốc, hàm Subset tìm tất cả các ứng cử viên được chứa trong giao dịch t như sau: Nếu ta bắt đầu tại một lá, ta tìm những tập trong nút lá này được chứa trong giao dịch t và bổ sung các mối quan hệ với chúng đối với tập kết quả mong muốn. Nếu ta đang ở một nút trong và ta đến được nó bằng việc băm mục i, ta băm trên mỗi mục đi sau i trong t và áp dụng một cách đệ quy thủ tục đó đối với nút này trong Bucket tương ứng. Đối với nút gốc, ta băm theo mỗi mục trong t.

Để thấy được tại sao hàm Subset trả lại tập các tham khảo mong muốn hãy để ý đến những gì sẽ xảy ra tại nút gốc. Đối với bất kỳ tập c nào được chứa trong giao dịch t, mục đầu tiên cần phải có trong t. Tại nút gốc, việc băm mọi mục trong t đảm bảo được rằng ta chỉ không biết các tập mà nó bắt đầu với một mục không nằm trong t. Những lí luận tương tự áp dụng cho các mức sâu hơn. Vì các mục trong bất kì tập nào cũng được sắp thứ tự, nếu ta đến được một nút hiện tại bằng việc băm mục i, ta chỉ cần quan tâm đến những mục trong t nó xuất hiện sau i.

// Bước tỉa: Xoá bớt tất cả các tập mục c ∈ Ck mà (k-1) tập con của c không phụ thuộc Lk-1.

1. for (∀ tập mục c ∈ Ck)

2. for (∀ (k-1) – tập con s của c)

3. if(s ∉ Lk-1)

4. delete c khỏi Ck;

**Nhận xét:** Thuật toán Apriori với n là độ dài lớn nhất của tập được sinh ra. Vậy thì thuật toán sẽ thực hiện duyệt toàn bộ các giao tác n+1 lần. Như vậy, nếu bỏ qua thời gian so sánh tìm sự xuất hiện của một mẫu trong một giao tác thì độ phức tạp của thuật toán Apriori là O(A) > O(n\*L) trong đó L là kích thước CSDL còn n là độ dài cần đạt được của các mẫu.

Ngoài ra, nếu độ hỗ trợ tối thiểu minsup bị thay đổi thì thuật toán sẽ phải thực hiện lại từ đầu, điều này sẽ rất mất thời gian. Thuật toán Apriori được xây dựng nhằm phát hiện các luật kết hợp giữa các đối tượng với độ hỗ trợ và độ tin cậy tối thiểu. [3]

# GIỚI THIỆU VỀ CÁC CÔNG CỤ

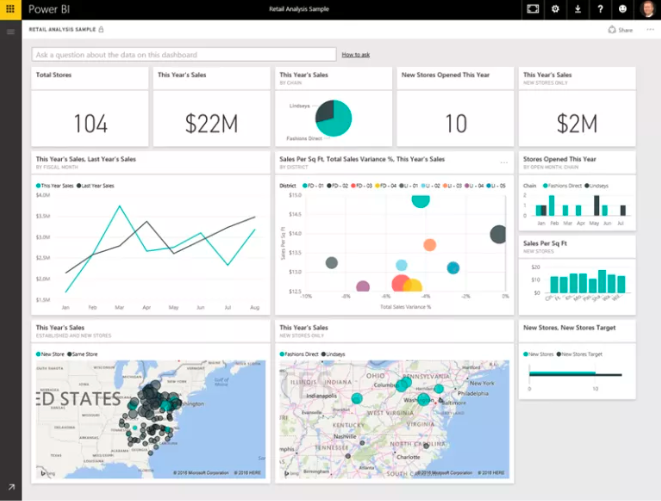
## Tổng quan về công cụ trực quan hóa Power BI

### Giới thiệu về Power BI

Power BI là một giải pháp phân tích kinh doanh cho phép bạn trực quan hóa dữ liệu của mình và chia sẻ thông tin chi tiết về tổ chức của bạn hoặc nhúng chúng vào ứng dụng hoặc trang web của bạn. Power BI kết nối với hàng trăm nguồn dữ liệu và làm cho dữ liệu của bạn trở nên sống động với bảng điều khiển và báo cáo trực tiếp. [4]

Power BI sẽ giúp bạn đưa ra được những phân tích dữ liệu và ra quyết định nhanh chóng, thuận lợi và dễ dàng. Power BI là một sản phẩm của gã khổng lồ Microsoft.

Power BI được công nhận là người dẫn đầu trong các nền tảng Analytics và Enterprise BI. Bên cạnh đó Power BI đã được Gartner đã công nhận Microsoft là nhà lãnh đạo về Power BI thể là công cụ báo cáo và trực quan cá nhân của bạn, đồng thời cũng đóng vai trò là công cụ phân tích và quyết định cho các dự án nhóm, bộ phận hoặc toàn bộ tập đoàn.



Hình . Môi trường làm việc của Power BI

### Các bộ phận của Power BI

* **Power BI Desktop:** Là một ứng dụng dùng trên máy tính để bàn.
* **Power BI Service:** Dịch vụ SaaS hỗ trợ trực tuyến
* **Power BI mobile:** Là phiên bản dùng trên điện thoại di động thông minh có hệ điều hành iOS, Android

### Công dụng của Power BI

Power BI giúp đưa ra quyết định nhanh chóng. Kết nối, mô hình hóa và sau đó khám phá dữ liệu của bạn với các báo cáo trực quan mà bạn có thể cộng tác, xuất bản và chia sẻ. Power BI tích hợp với các công cụ khác, bao gồm Microsoft Excel, do đó bạn có thể tăng tốc nhanh chóng và làm việc liền mạch với các giải pháp hiện có của mình.

* **Khởi tạo:** Kết nối với dữ liệu của bạn bất cứ nơi nào. Sau đó khám phá dữ liệu của bạn với hình ảnh tương tác tuyệt đẹp.
* **Hợp tác và chia sẻ:** Xuất bản báo cáo và bảng điều khiển, cộng tác với nhóm của bạn và chia sẻ những hiểu biết bên trong và bên ngoài tổ chức của bạn làm việc.
* **Truy cập thông tin ở bất cứ nơi đâu:** Bạn có thể dễ dàng xem thông tin ở cả máy tính và ở chiếc điện thoại thông mình có cài Power Bi mobile.

Như vậy bạn có thể sử dụng Power BI để:

* Tạo ra các báo cáo có tính tương tác caoPower BI cung cấp cho bạn các công cụ để chuyển đổi, phân tích và hiển thị dữ liệu.
* Phân tích mức độ ổn địnhVới Power BI, bạn rất dễ dàng để tạo ra những mô hình tái sử dụng dữ liệu (reuseable data) để đảm bảo sự nhất quán giữa các báo cáo,  tiến hành các dự đoán, đưa ra các xu thế trong tương lai dựa vào các dữ liệu hiện tại và trong quá khứ.
* Nhúng báo cáo vào các ứng dụng của bạnTích hợp ứng dụng của bạn với Power BI, các báo cáo tương tác và biểu đồ thời gian thực sẽ dễ dàng được chia sẻ đến những người dùng khác hay khách hàng của bạn.

## Tổng quan về Microsoft SQL Server 2014

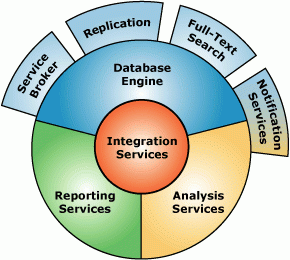
### Giới Thiệu SQL Server

SQL Server là một hệ quản trị cơ sở dữ liệu quan hệ (Relational Database Management System (RDBMS) ) sử dụng câu lệnh SQL (Transact-SQL) để trao đổi dữ liệu giữa máy Client và máy cài SQL Server. Một RDBMS bao gồm databases, database engine và các ứng dụng dùng để quản lý dữ liệu và các bộ phận khác nhau trong RDBMS.

SQL Server được tối ưu để có thể chạy trên môi trường cơ sở dữ liệu rất lớn (Very Large Database Environment) lên đến Tera-Byte và có thể phục vụ cùng lúc cho hàng ngàn user. SQL Server có thể kết hợp “ăn ý” với các server khác như Microsoft Internet Information Server (IIS), E-Commerce Server, Proxy Server….

### Các thành cơ bản trong SQL Server

SQL Server được cấu tạo bởi nhiều thành phần như Database Engine, Reporting Services, Notification Services, Integration Services, Full Text Search Service…. Các thành phần này khi phối hợp với nhau tạo thành một giải pháp hoàn chỉnh giúp cho việc lưu trữ và phân tích dữ liệu một cách dễ dàng.



Hình . Các thành phần cơ bản trong SQL

**Database Engine**

Cái lõi của SQL Server

Ðây là một engine có khả năng chứa data ở các quy mô khác nhau dưới dạng table và support tất cả các kiểu kết nối (data connection) thông dụng của Microsoft như

ActiveX Data Objects (ADO), OLE DB, and Open Database Connectivity (ODBC).

Ngoài ra nó còn có khả năng tự điều chỉnh (tune up) ví dụ như sử dụng thêm các tài nguyên (resource) của máy khi cần và trả lại tài nguyên cho hệ điều hành khi một user log off.

**Replication**

– Cơ chế tạo bản sao

**Integration Services (DTS)**

– Integration Services là một tập hợp các công cụ đồ họa và các đối tượng lập trình cho việc di chuyển, sao chép và chuyển đổi dữ liệu.

**Analysis Services**

– Một dịch vụ phân tích dữ liệu rất hay của Microsoft

Dữ liệu (Data) chứa trong database sẽ chẳng có ý nghĩa gì nhiều nếu như bạn không thể lấy được những thông tin (Information) bổ ích từ đó. Do đó Microsoft cung cấp cho bạn một công cụ rất mạnh giúp cho việc phân tích dữ liệu trở nên dễ dàng và hiệu quả bằng cách dùng khái niệm hình khối nhiều chiều (multi-dimension cubes) và kỹ thuật “khai phá dữ liệu” (data mining).

**Notification Services**

Dịch vụ thông báo Notification Services là nền tảng cho sự phát triển và triển khai các ứng dụng tạo và gửi thông báo. Notification Services có thể gửi thông báo theo địch thời đến hàng ngàn người đăng ký sử dụng nhiều loại thiết bị khác nhau.

**Reporting  Services**

Reporting Services bao gồm các thành phần server và client cho việc tạo, quản lý và triển khai các báo cáo. Reporting Services cũng là nền tảng cho việc phát triển và xây dựng các ứng dụng báo cáo.

**Full Text Search Service**

Dịch vụ SQL Server Full Text Search là một dịch vụ đặc biệt cho đánh chỉ mục và truy vấn cho dữ liệu văn bản không cấu trúc được lưu trữ trong các CSDL SQL Server. Đánh chỉ mục với Full Text Search có thể dduwowcj tạo trên bất kỳ cột dựa trên dữ liệu văn bản. Nó sẽ rất hiệu quả cho việc tìm các sử dụng toán tử LIKE trong SQL với trường hợp tìm văn bản.

**Service Broker**

Được sử dụng bên trong mỗi Instance, là môi trường lập trình cho việc các ứng dụng nhảy qua các Instance. Service Broker giao tiếp qua giao thức TCP/IP và cho phép các component khác nhau có thể được đồng bộ cùng nhau theo hướng trao đổi các message. Service Broker chạy như một phần của bộ máy cơ sở dữ liệu, cung cấp một nền tảng truyền message tin cậy và theo hàng đợi cho các ứng dụng SQL Server. [4]

## Microsoft SQL Server Analysis Services (SSAS)

Analysis Services là môt dịch vụ hỗ trợ mạnh mẽ việc phân tích, khai thác thông tin tiềm tàng bên trong môt hệ quản tri cơ sở dữ liệu dựa vào khái niệm khối đa chiều(multiple-dimension cube) và khai phá dữ liệu.

Analysis Services là công cụ dễ sử dụng, tích hợp và linh động giúp định nghĩa các chiều, xây dựng các khối dữ liệu đa chiều và cung cấp các chương trình ứng dụng truy xuất tới các khối này.

Microsoft SQL Server Analysis Services (SSAS) đáp ứng nhu cầu xử lý phân tích dữ liệu OLAP và các ứng dụng khai phá dữ liệu. Analysis Services hỗ trợ OLAP bằng cách cho phép bạn thiết kế, tạo ra, và quản lý các cấu trúc đa chiều có chứa dữ liệu được tổng hợp từ các nguồn dữ liệu khác, chẳng hạn như cơ sở dữ liệu quan hệ. Đối với các ứng dụng khai thác dữ liệu Analysis Services cho phép bạn thiết kế, tạo ra, và hình dung mô hình khai thác dữ liệu được xây dựng từ các nguồn dữ liệu khác bằng cách sử dụng một loạt các tiêu chuẩn các thuật toán khai thác dữ liệu.

SSAS cho phép các ứng dụng hoạt động theo chế độ Client/Server. Thành phần server được triển khai như một dịch vụ của Microsoft Windows bao gồm một số thành phần thực hiện các chức năng:

* Phân tích câu lệnh nhận được từ client.
* Quản lý metadata.
* Xử lý tính toán.
* Lưu trữ dữ liệu.
* Tạo các khối tích hợp.
* Lập lịch truy vấn.
* Quản lý tài nguyên.

SSAS hỗ trợ nhiều dịch vụ trên cùng một máy chủ, mỗi dịch vụ là một đối tượng riêng biệt. Thành phần client truyền thông với server bằng giao thức XML for Analysis (XML/A) thông qua cơ chế truy vấn và trả lời tương tự như web service. Câu lệnh truy vấn sử dụng ngôn ngữ SQL, MDX (Multidimensional Expressions), DMX (Data Mining Extensions). Vài provider cũng được SSAS cung cấp để hỗ trợ các ngôn ngữ lập trình khác nhau.

## Môi trường triển khai ứng dụng (BIDS)

Business Intelligence Development Studio (BIDS) là môi trường phát triển ứng dụng SSAS. Môi trường này được tích hợp với Microsoft Analysis Service. Phần lớn thời gian làm việc với SQL Server Data Mining là làm việc với BIDS. Nó là công cụ cho phép tổ chức và khai thác kho dữ liệu (xử lý phân tích trực tuyến) cũng như xây dựng mô hình khai phá dữ liệu rất dễ sử dụng và hiệu quả của Microsoft.

BIDS có hai chế độ hoạt động: Offline và Immediate.

Làm việc với chế độ Immediate, người dùng kết nối trực tiếp và liên tục đến Analysis Service server. Khi tạo mới hoặc thay đổi các đối tượng đã có, các hoạt động này đều được thay đổi trên server. Mặc dù chế độ làm việc này rất trực quan và dễ hiểu. Tuy nhiên, nó cũng có nhiều hạn chế như các thao tác sai không thể phục hồi, các tác động không mong muốn từ các phiên làm việc khác, tính nhất quán giữa các ứng dụng sử dụng chung dữ liệu. Ngoài ra chế độ này chỉ bó hẹp trong một sơ sở dữ liệu nhất định.

# PHÂN TÍCH PHẢN HỒI VỀ VIỆC SỬ DỤNG DỊCH VỤ TẠI CHUỖI RỬA XE THÔNG MINH VIETWASH

## Thực trạng và yêu cầu

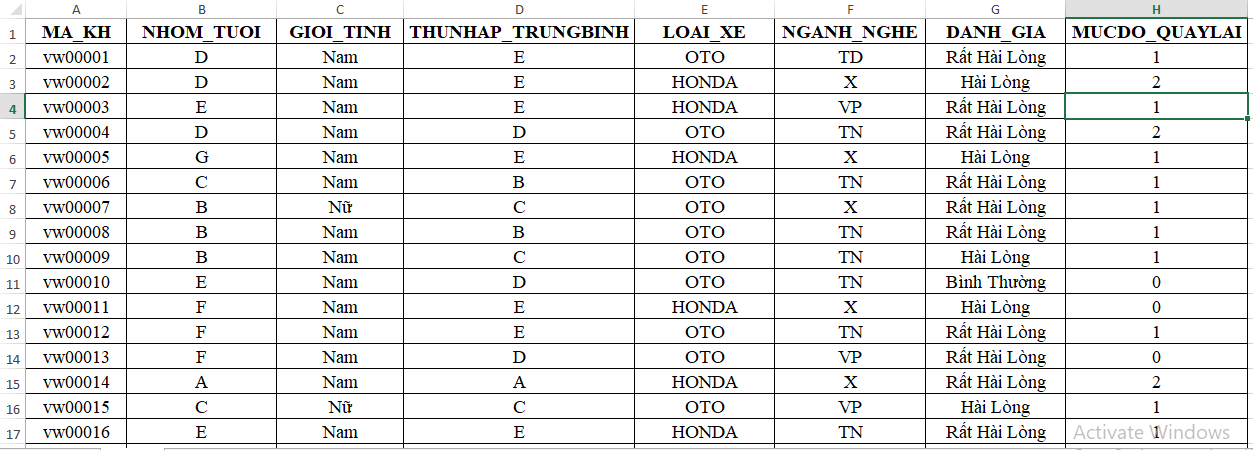
Chuỗi rửa xe VIETWASH đã tiến hành khảo sát mức độ hài lòng của khách hàng về dịch vụ mà công ty đã cung cấp. Dữ liệu thu được gồm 8 cột: Mã khách hàng, Nhóm tuổi, Giới tính, Thu nhập trung bình, Loại xe, Ngành nghề, Đánh giá, Mức độ quay lại và 5000 dòng.

Công ty muốn khai thác dữ liệu trên để đánh giá phân tích mức độ hài lòng của khách hàng đối với dịch vụ mà công ty cung cấp. Vì vậy, bài báo cáo này sẽ sử dụng dữ liệu mà công ty đã thu thập cùng với các công cụ: Power BI, Microsoft SQL Server 2014 và ứng dụng luật kết hợp để xây dựng mô hình khai phá dữ liệu. Qua đó, phân tích mức độ hài lòng của khách hàng để quản lý cũng như hình thành chính sách chăm sóc hợp lý nhằm thỏa mãn nhu cầu cũng như gia tăng được mức độ quay lại của khách hàng.

## Đề xuất qui trình

Hình . Đề xuất quy trinhg khai phá dữ liệu

### Thu thu thập dữ liệu



Hình . Thu thập dữ liệu

Dữ liệu thu được 5000 dòng và 8 cột gồm : Mã KH, Nhóm tuổi, Giới tính, Thu nhập trung bình, Loại xe, Ngành nghề, Đánh giá, Mức độ quay lại.

* **Nhóm tuổi** gồm 7 nhóm :

A: < 20t (nhỏ hơn 20t)

B: 20 - 25t

C: 25 - 30t

D: 30 - 35t

E: 35 - 40t

F: 40 - 45t

G: > 45t (trên 45 tuổi)

* **Giới tính** gồm:

Nam

Nữ

* **Thu nhập trung bình** gồm 5 mức:

A < 5tr (nhỏ hơn 5tr)

B: 5-10tr

C: 10 - 20tr

D: 20 - 30tr

E: > 30tr (lớn hơn 30tr)

* **Loại xe** gồm 2 loại:

HONDA

OTO

* **Ngành nghề** được chia làm 4 loại :

TD: Tự do

X: Không xác định

VP: Văn phòng

TN: Thương nhân

* **Đánh giá** gồm 5 mức:

Rất Hài Lòng

Hài Lòng

Bình Thường

Không Hài Lòng

Rất Không Hài Lòng

* **Mức độ quay lại** gồm:

1 lần

2 lần

3 lần

4 lần

5 lần

6 lần

### Thiết kế dữ liệu

Bảng dữ liệu được thiết kế có kiểu dữ liệu và kích thước như sau:

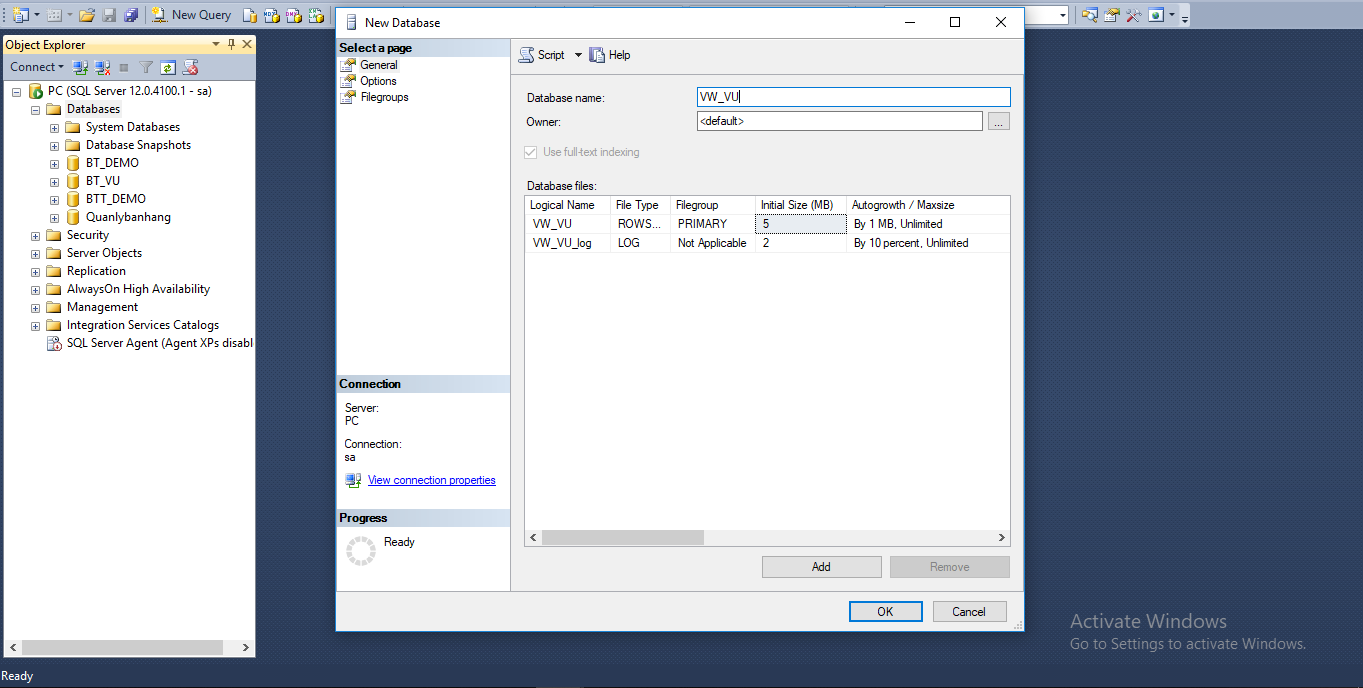
|  |  |  |  |
| --- | --- | --- | --- |
| **Tên trường** | **Kiểu dữ liệu** | **Kích thước** | **Mô tả** |
| MA\_KH | nvarchar | 7 | Mã khách hàng |
| NHOM\_TUOI | nvarchar | 1 | Nhóm tuổi |
| GIOI\_TINH | nvarchar | 3 | Giới tính |
| THUNHAP\_TRUNGBINH | nvarchar | 1 | Thu nhập trung bình |
| LOAI\_XE | nvarchar | 5 | Loại xe |
| NGANH\_NGHE | nvarchar | 2 | Ngành nghề |
| DANH\_GIA | nvarchar | 30 | Đánh giá |
| MUCDO\_QUAYLAI | int |  | Mức độ quay lại |

Bảng Kiểu dữ liệu và kích thước

### Cài đặt dữ liệu vào hệ thống

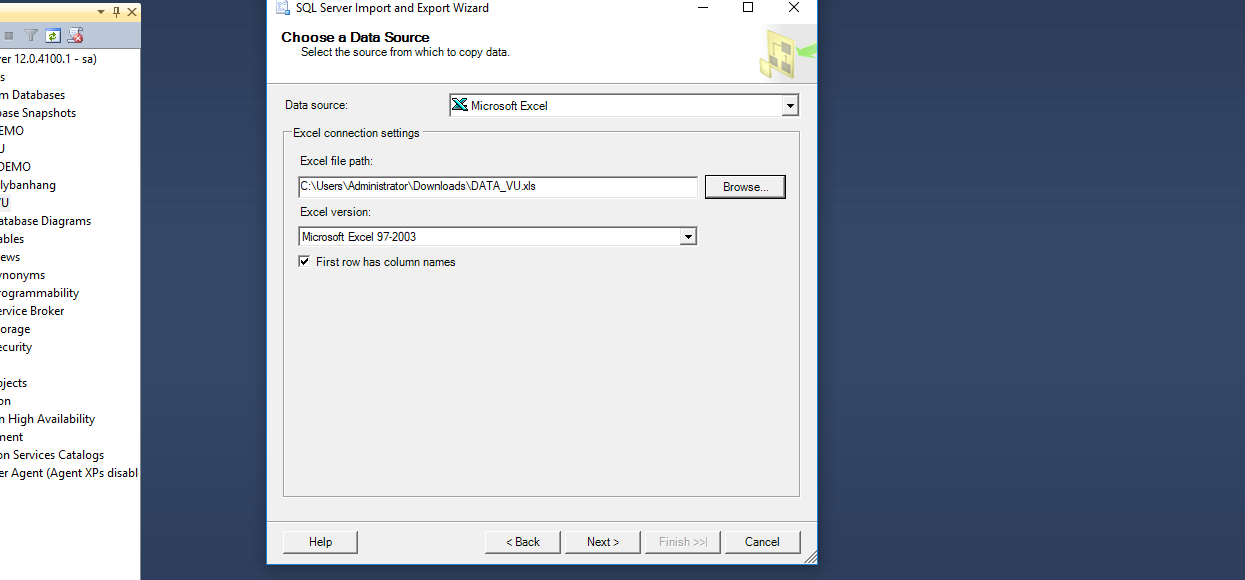
Để cài đặt dữ liệu vào hệ thống bước đầu tiên ta cần tạo một cơ sở dữ liệu mới.

Vào Database 🡺 New Database 🡺 Database Name: VW\_VU.

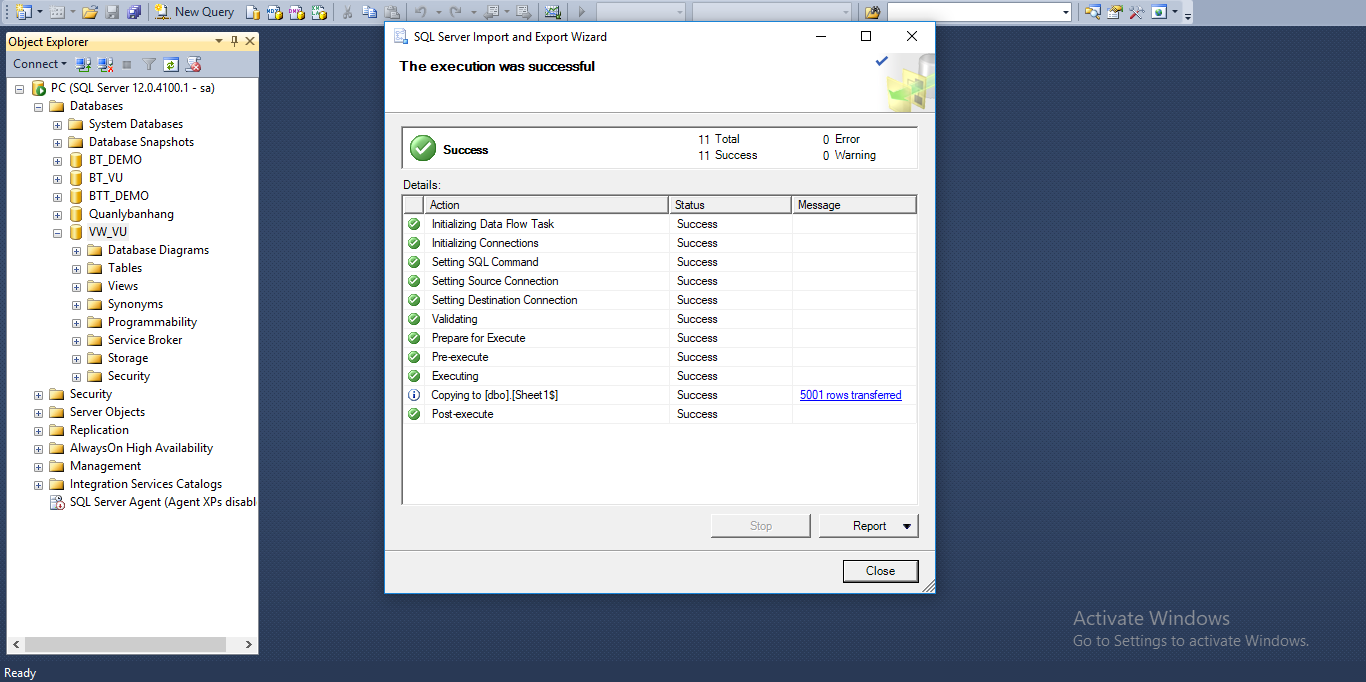


Hình . Tạo Data base mới trong Microsoft SQL

Tiếp theo Import dữ liệu vào Microsoft SQL Server 2014: Kích phải chuột vào database VW\_VU🡺 Task 🡺 Import Data 🡺 Next.

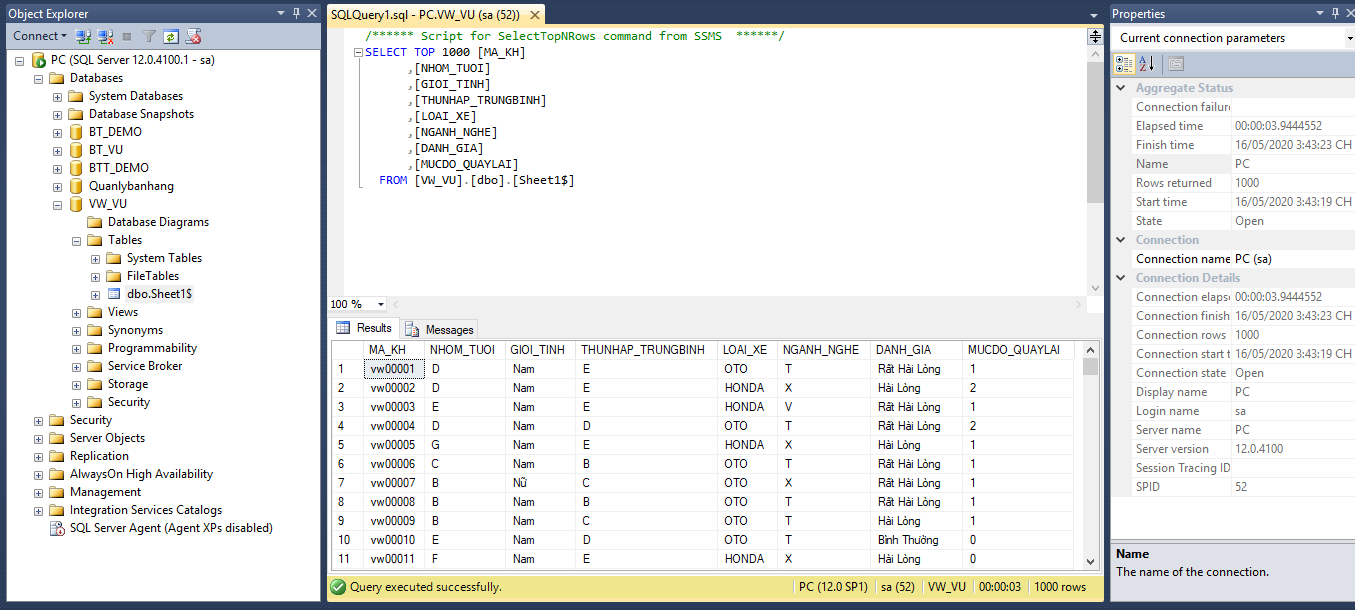


Hình . Import dữ liệu vào Database



Hình . Import dữ liệu thành công

Dữ liệu đã được copy vào bảng dbo.Sheet1$

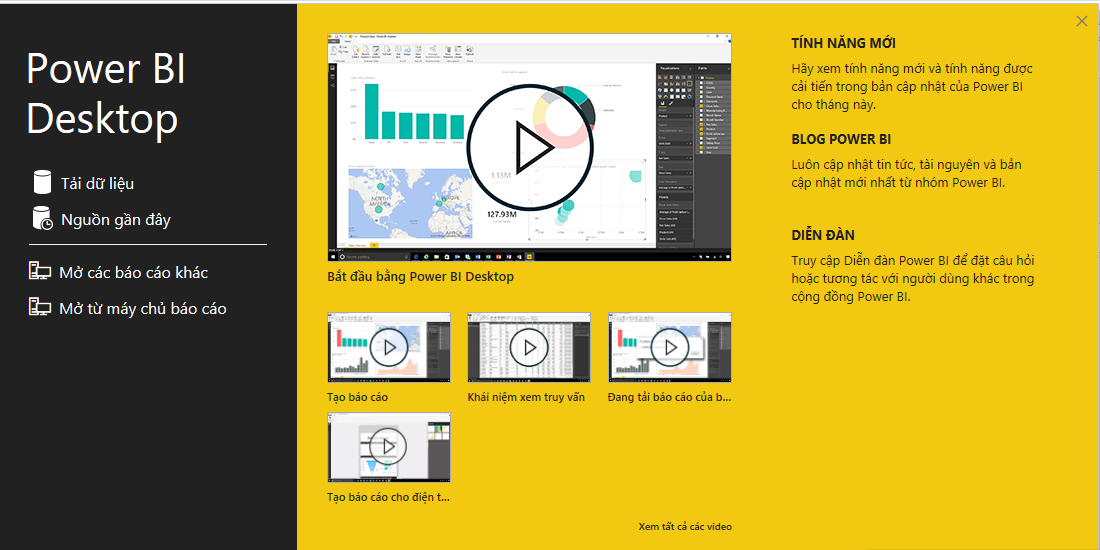


Hình . Dữ liệu trong Database

### Trực quan hóa dữ liệu

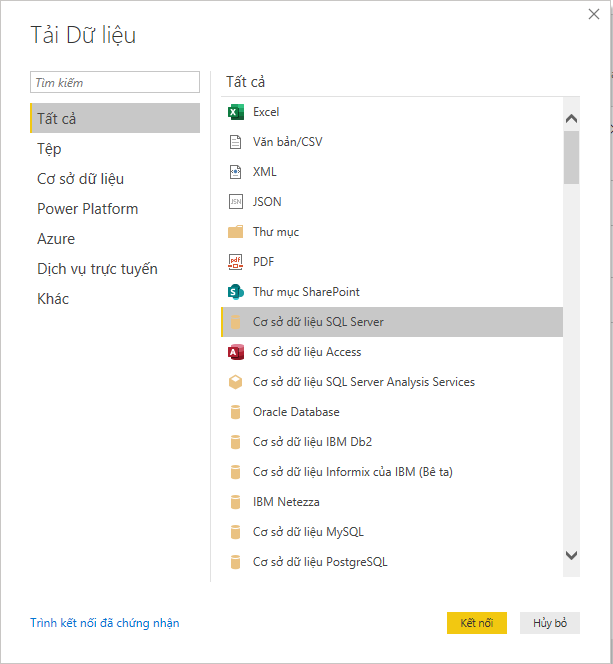
Một trong những công cụ trực quan hóa dữ liệu phổ biến nhất hiện nay là Power BI cho nên trong phạm vi đề tài nay sử dụng Power BI để trực quan hóa dữ liệu.

Các bước sử dụng Power BI như sau:



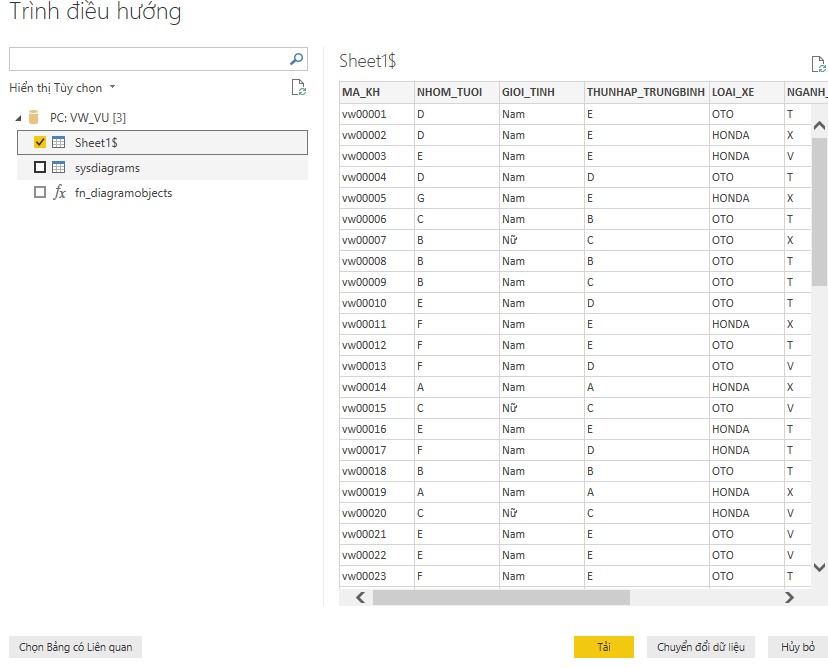
Hình . Sử dụng công cụ Power BI

Đầu tiên chúng ta cần tải dữ liệu vào, nguồn dữ liệu từ Excel. Để thêm dữ liệu vào Power BI: Tải dữ liệu 🡺SQLSever 🡺 Kết nối



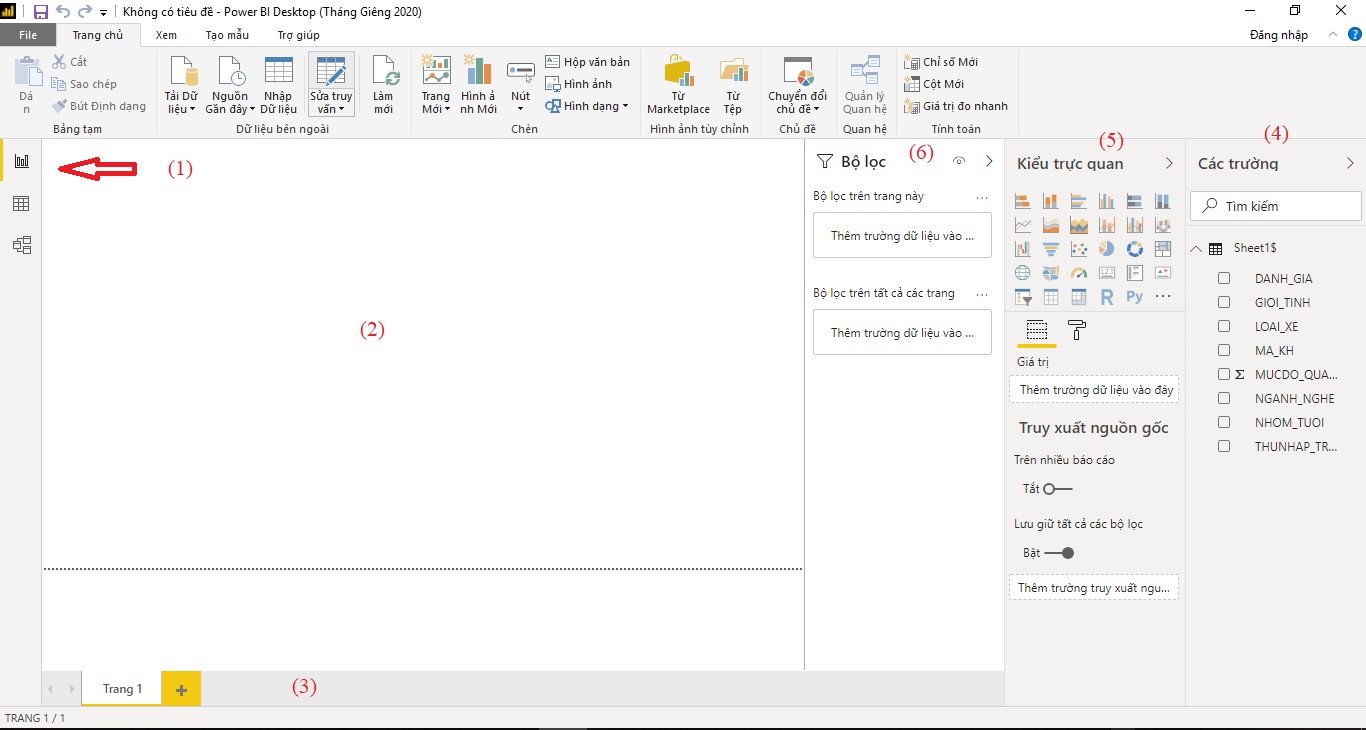
Hình 4.8 Tải dữ liệu vào Power BI

Trình điều hướng sau khi tải dữ liệu



Hình . Trình điều hướng sau khi tải dữ liệu

Màn hình làm việc của Power BI sau khi tải dữ liệu



Hình . Màn hình làm việc sau khi tải dữ liệu

Giao diện xử lý sẽ có 6 mục mà bạn cần chú ý như trên, chính là các khu vực thường xuyên được sử dụng:

1. Nơi lựa chọn xem báo cáo ở dạng biểu đồ hay các con số và mối quan hệ giữa các cột dữ liệu.
2. Vùng hiển thị, nơi chứa biểu đồ để bạn xử lý và sắp xếp.
3. Cũng giống như nhiều công cụ khác Power BI cho phép tạo ra nhiều trang. Để bạn tùy ý sử dụng cùng dữ liệu cho nhiều mục đích khác nhau.
4. Fields: nơi chứa các bảng dữ liệu và các cột được dùng để vẽ biểu đồ.
5. Visualizations: khu vực chỉnh sửa biểu đồ, nơi bạn chọn loại biểu đồ, màu sắc, cách hiển thị của từng dữ liệu…
6. Filters: cho phép thêm vào các điều kiện để xử lý dữ liệu trong biểu đồ. Ví dụ: hiển thị Top 10 quốc gia đứng đầu trong báo cáo.

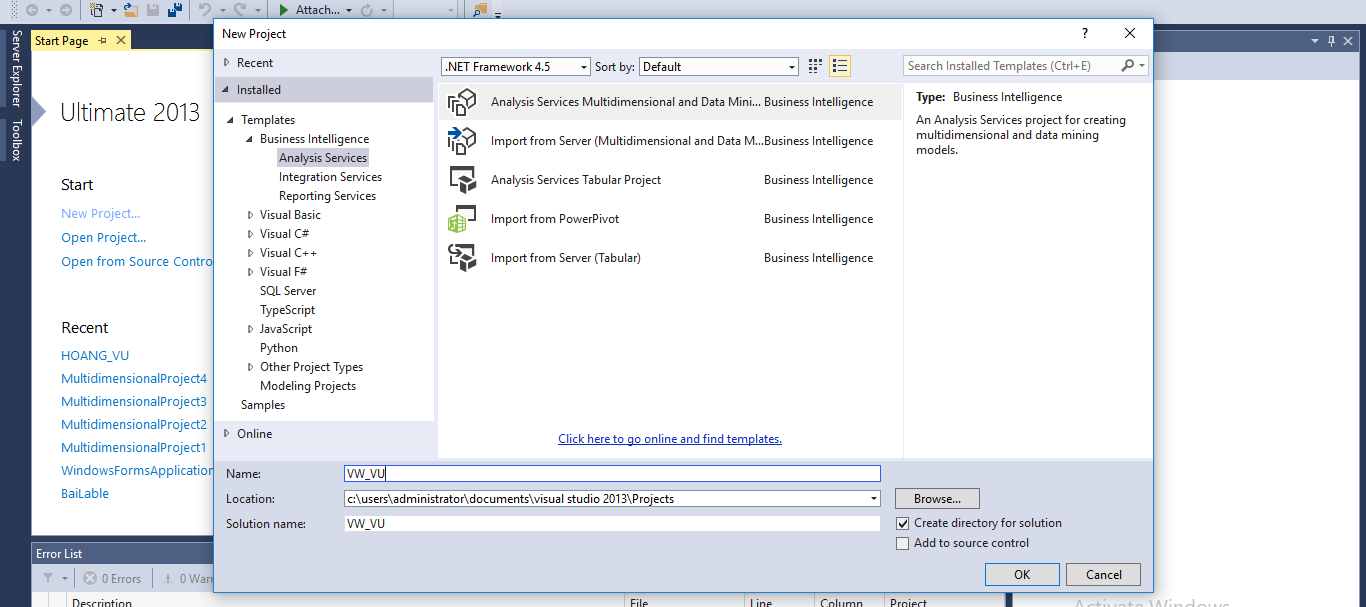
Kết quả phân tích được để ở chương cuối.

### Xây dựng mô hình khai phá dữ liệu trên BIDS

Tiếp theo chúng ta sẽ xây dựng mô khai phá dữ liệu trên BIDS, các bước xây dựng mô hình như sau:

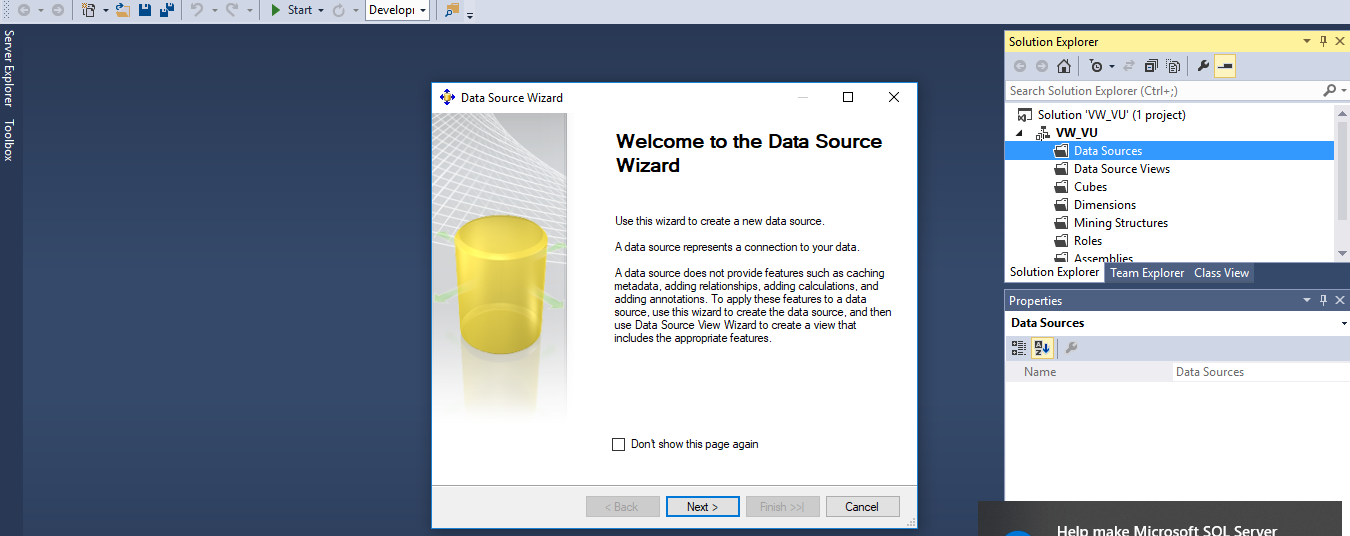
* **Tạo Data Source**

Tạo project mới có tên là : VW\_VU



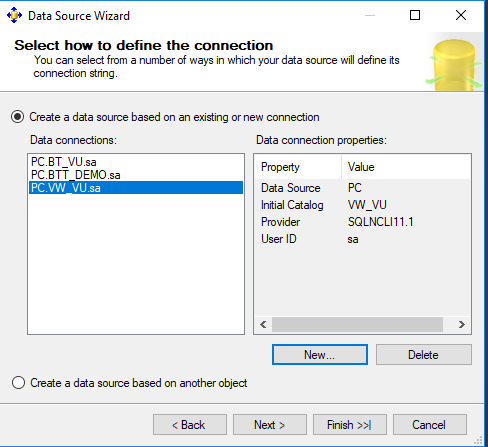
Hình . Tạo project mới

Tạo Data Source bằng cách kích chuột phải Data Source 🡺 New Data Source 🡺 Xuất hiện hộp thoại Welcome to the Data Soure Wizard 🡺 Chon Next.



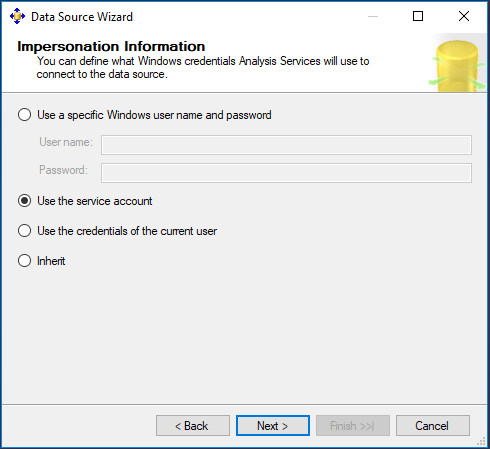
Hình . Tạo Data Source Wizard

Tại hộp thoại Select how to define the connection , kiểm tra check box Create a data source based on an existing or new conection rồi nhấn Next.



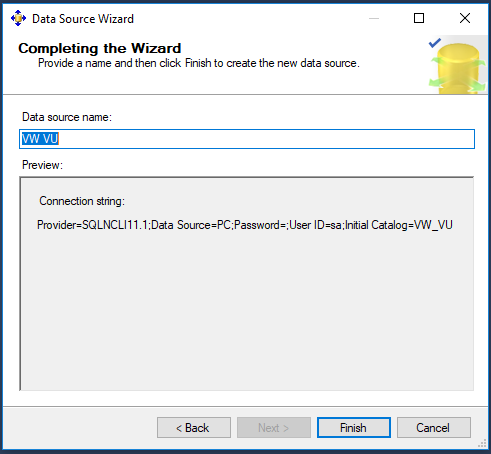
Hình . Tạo Data Source Wizard

Tại cửa sổ Impersonation Information chọn Use the service account và nhấn Next.



Hình . Tạo Data Source Wizard

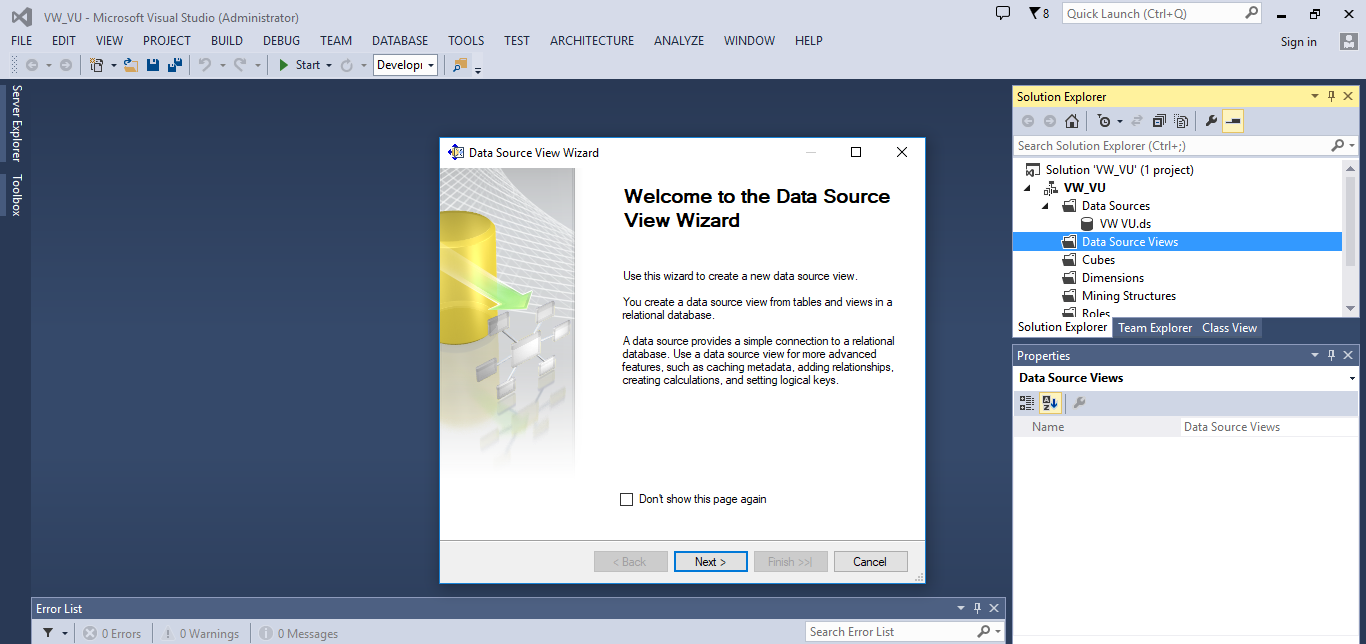
Tại màn hình Completing the wizard, thiết lập tên của Data Source là BTT\_TRUC và chọn Finish.



Hình . Tạo Data Source Wizard

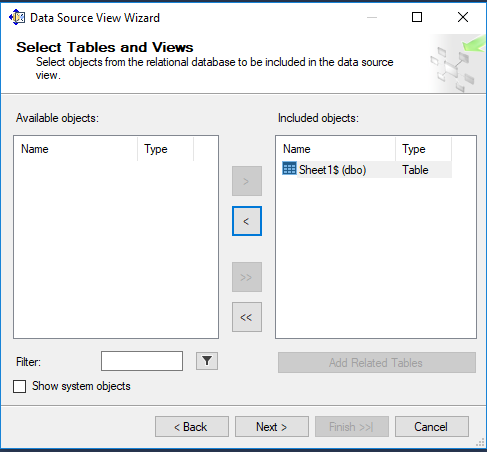
* **Tạo Data Source View**

Kích phải chuột Data Source View, chọn New Data Source Views 🡺 Xuất hiện cửa sổ Welcome to the Data Source View Wizard 🡺 Chọn Next.



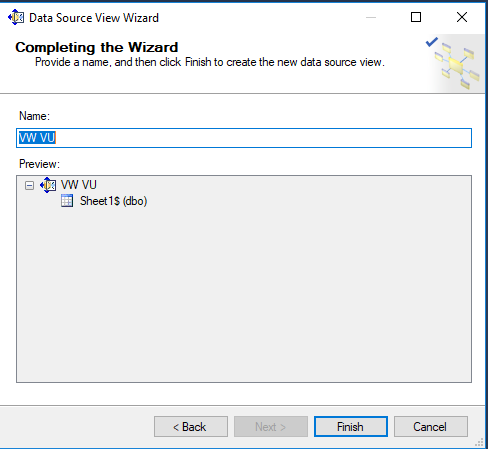
Hình . Tạo Data Source View

Tại hộp thoại Select Tables and Views, chọn bảng Sheet1 để thêm bảng này  
vào danh sách Included objects, rồi nhấn Next.



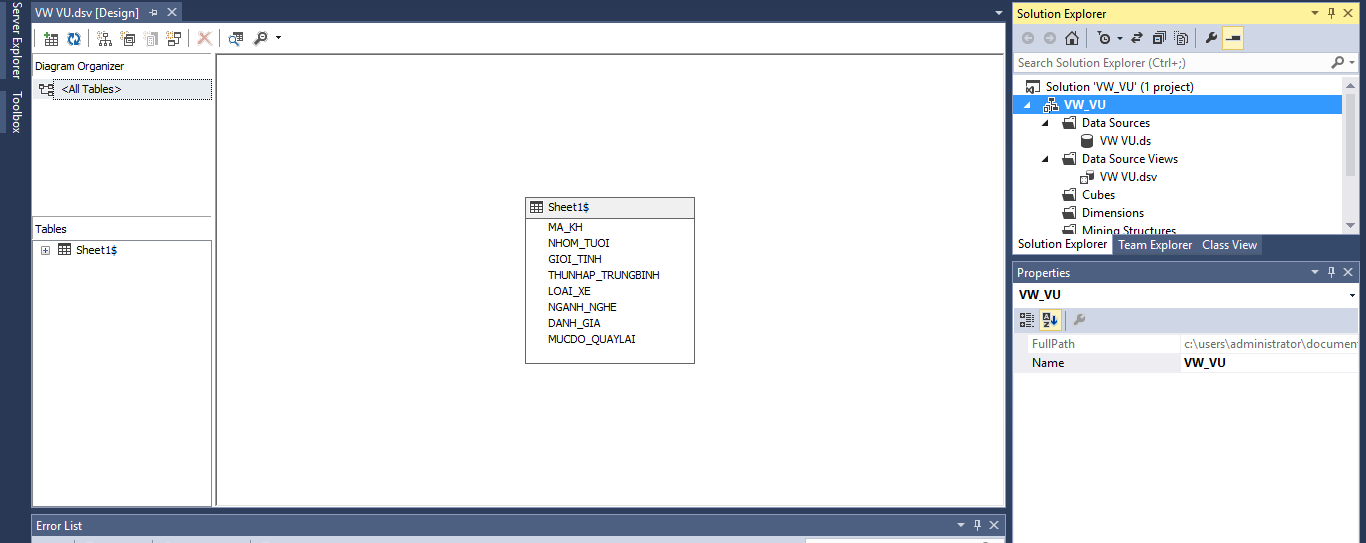
Hình . Tạo Data Source View

Tại màn hình Completing the Wizard thiết lập Name là BTVU 🡺 Nhấn Finish.



Hình . Tạo Data Source View

Sau thực hiện xong bước này ta được khung hình về bảng dữ liệu như bên dưới.



Hình . Tạo Data Source View thành công

### Ứng dụng luật kết hợp phân tích phản hồi khách

***Quy trình xây dựng mô hình luật kết hợp.***

Các bước thực thi luật kết hợp:

Hình . Các bước thực thi luật kết hợp

* **Tạo data minning structures AR\_model**

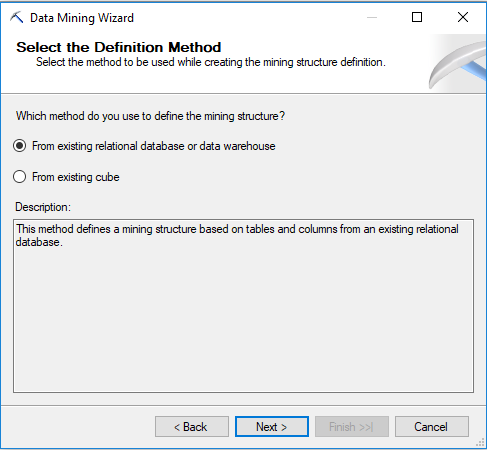
Kích phải chuột vào Mining Structures 🡺 chọn New Mining Structure.



Hình . Giao diện tạo Mining Structure

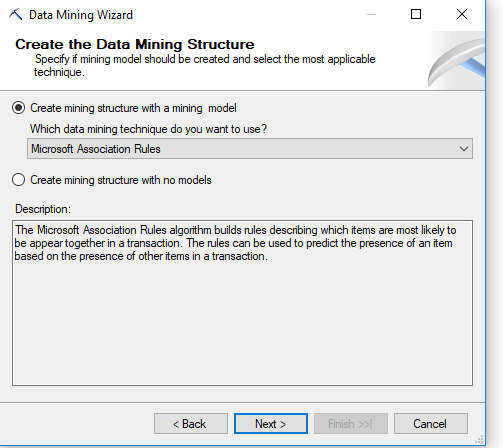
Data Mining Wizard sẽ tạo ra cấu trúc và mô hình khai phá dữ liệu. Sau đó chọn phương pháp để xử lý mô hình dựa vào các bảng và các cột từ cơ sở dữ liệu quan hệ, chọn From existing relational database or data warehouse.

Ta tiếp tục chọn Next.



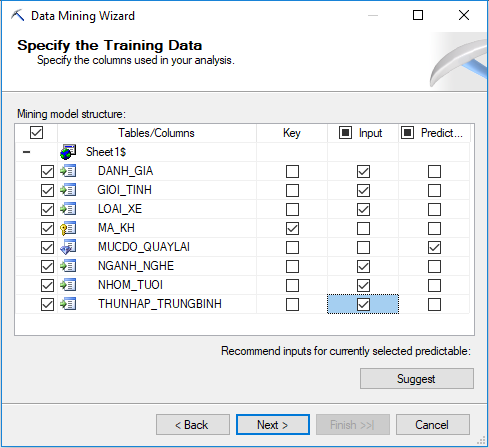
Hình . Tạo Mining Structure

Ở hộp thoại Create the Data Mining Structure tích vào ô Create mining structure with a mining model và chọn thuật toán Microsoft Association Rules 🡺 Chọn Next.



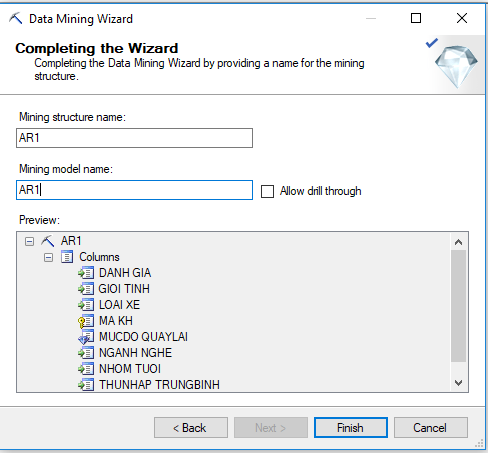
Hình . Lựa chọn mô hình luật kết hợp

Chọn các thuộc tính để training. Chọn khóa là MA\_KH, Chọn các thuộc tính input là: DANH\_GIA, GIOI\_TINH, LOAI\_XE, NGANH\_NGHE, NHOM\_TUOI, THUNHA\_TRUNGBINH. Chọn thuộc tính Predict là MUCDO\_QUAYLAI.



Hình . Chọn thuộc tính để training (Mô hình AR1)

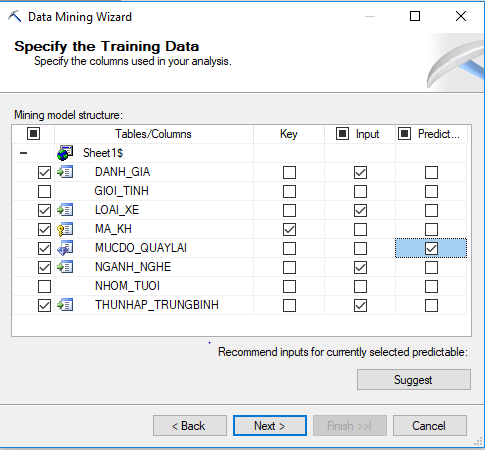
Đặt tên cho mô hình là AR1 sau đó chọn Finish.



Hình . Đặt tên mô hình luật kết hợp (Mô hình AR1)

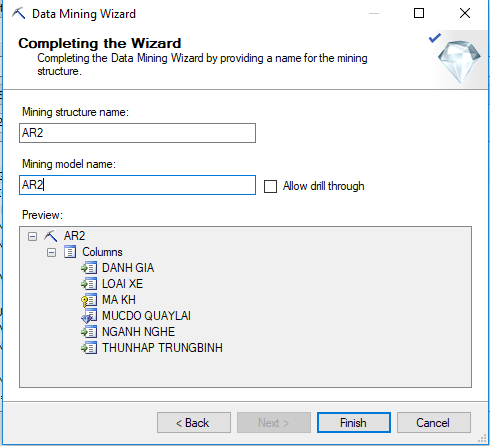
Chọn các thuộc tính để training

Chọn khóa là MA\_KH. Thuộc tính input là: LOAI\_XE, NGANH\_NGHE, DANH\_GIA, THUNHAP\_TRUNGBINH. Thuộc tính Predict là MUCDO\_QUAYLAI. Loại bỏ thuộc tính GIOI\_TINH và NHOM\_TUOI.



Hình . Chọn thuộc tính để Training (Mô hình AR2)

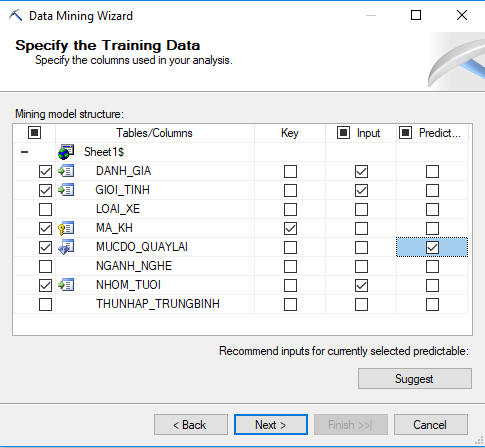
Đặt tên mô hình là AR2 và chọn Finish.



Hình . Đặt tên mô hình luật kết hợp (Mô hình AR2)

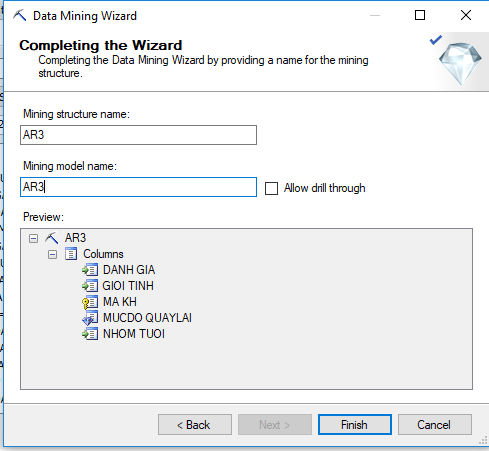
Chọn các thuộc tính để training

Chọn khóa là MA\_KH. Thuộc tính input là: DANH\_GIA, GIOI\_TINH, NHOM\_TUOI . Thuộc tính Predict là MUCDO\_QUAYLAI. Loại bỏ thuộc tính LOAI\_XE, NGANH\_NGHE, THUNHAP\_TRUNGBINH.



Hình . Chọn thuộc tính training (Mô hình AR3)

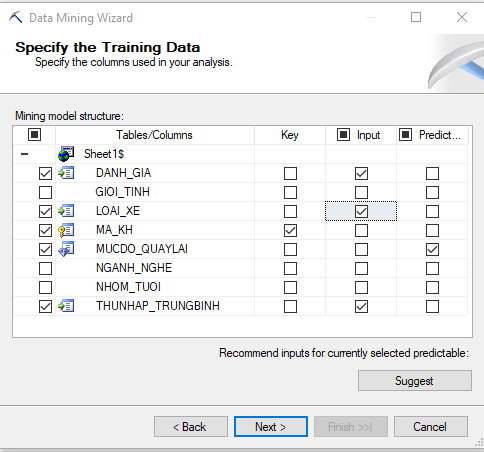
Đặt tên mô hình là AR3 và chọn Finish.



Hình . Đặt tên mô hình luật kết hợp (Mô hình AR3)

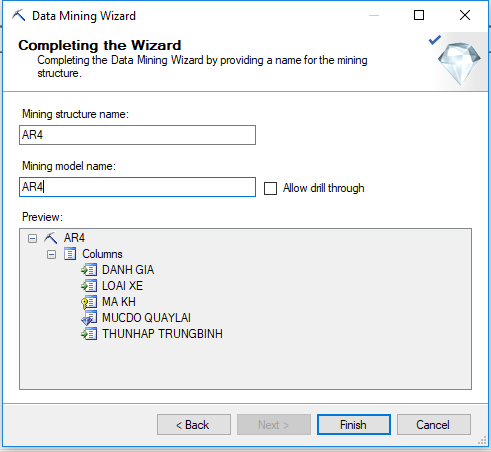
Chọn các thuộc tính để training

Chọn khóa là MA\_KH. Thuộc tính input là: DANH\_GIA, LOAI\_XE, THUNHAP\_TRUNGBINH . Thuộc tính Predict là MUCDO\_QUAYLAI. Loại bỏ thuộc tính GIOI\_TINH, NGANH\_NGHE, NHOM\_TUOI.



Hình . Chọn thuộc tính để training (Mô hình AR4)

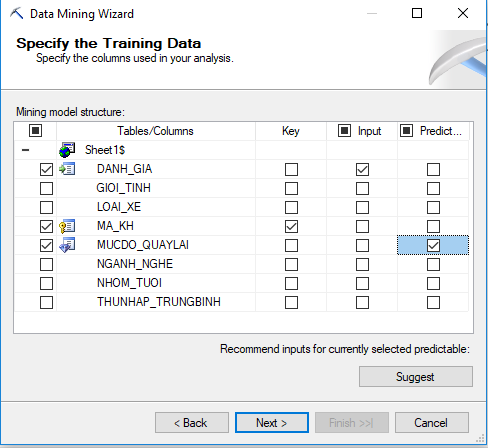
Đặt tên mô hình là AR4 và chọn Finish



Hình . Đặt tên mô hình luật kết hợp (Mô hình AR4)

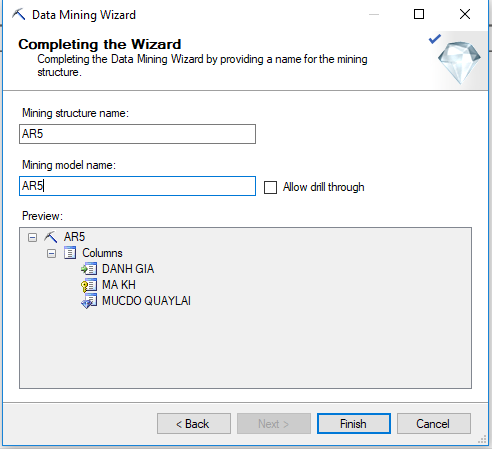
Chọn các thuộc tính để training

Chọn khóa là MA\_KH. Thuộc tính input là: DANH\_GIA. Thuộc tính Predict là MUCDO\_QUAYLAI. Loại bỏ thuộc tính GIOI\_TINH, LOAI\_XE, NGANH\_NGHE, THUNHAP\_TRUNGBINH, NHOM\_TUOI.



Hình . Chọn thuộc tính để training (Mô hình AR5)

Đặt tên mô hình là AR5 và chọn Finish.



Hình . Đặt tên mô hình luật kết hợp (Mô hình AR5)

***Hiệu chỉnh tham số cho mô hình***

**MAXIMUM\_ITEMSET\_COUNT**

Chỉ số lượng tối đa của các tập phổ biến được sinh ra. Nếu không được quy định  
số lượng cụ thể, giá trị mặc định sẽ được sử dụng.

Chọn mặc định là 20000.

**MAXIMUM\_ITEMSET\_SIZE**

Chỉ định số lượng tối đa các mục được phép trong tập phổ biến. Thiết lập giá trị này là 0 xác định rằng không có giới hạn kích thước của tập phổ biến.

Giá trị này có khả năng có thể giảm thời gian là cần thiết cho việc tạo ra các mô hình, bởi vì việc xử lý của mô hình dừng lại khi giới hạn được đạt đến.

Để mặc định là 3.

**MAXIMUM\_SUPPORT**

Chỉ định số lượng tối đa của các trường hợp mà tập phổ biến được hỗ trợ.Tham số này có thể được sử dụng để loại bỏ các mục mà xuất hiện thường xuyên.

Nếu giá trị này nhỏ hơn 1, tức là giá trị đại diện cho một tỷ lệ phần trăm của tổng số các trường hợp. Những giá trị lớn hơn 1 đại diện cho số lượng tuyệt đối của các trường hợp mà có thể chứa các tập phổ biến.

Mặc định là 1.

**MINIMUM\_ITEMSET\_SIZE**

Chỉ định số lượng tối thiểu các mục được phép trong tập phổ biến. Nếu tăng giá trị tham số này, mô hình có thể chứa các tập phổ biến ít hơn.

Mặc định là 1.

**MINIMUM\_PROBABILITY**

Xác định xác suất tối thiểu mà một luật sinh ra là đúng.

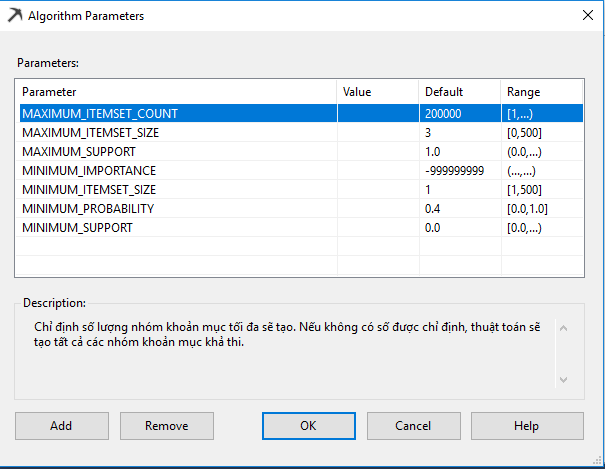
Mặc định 0.4.

**MINIMUM\_SUPPORT**

Chỉ định số lượng tối thiểu các trường hợp mà phải chứa các tập phổ biến trước khi thuật toán tạo ra các tập luật.

Nếu thiết đặt giá trị này nhỏ hơn 1, số lượng tối thiểu các trường hợp được tính như là một tỷ lệ phần trăm của tổng số các trường hợp. Nếu thiết đặt giá trị này là một số nguyên lớn hơn 1, xác định số lượng tối thiểu các trường hợp được tính như số các trường hợp mà phải có các tập phổ biến. Thuật toán có lẽ tự động tăng giá trị của các tham số này nếu bộ nhớ được giới hạn.

Mặc định là 0.0.

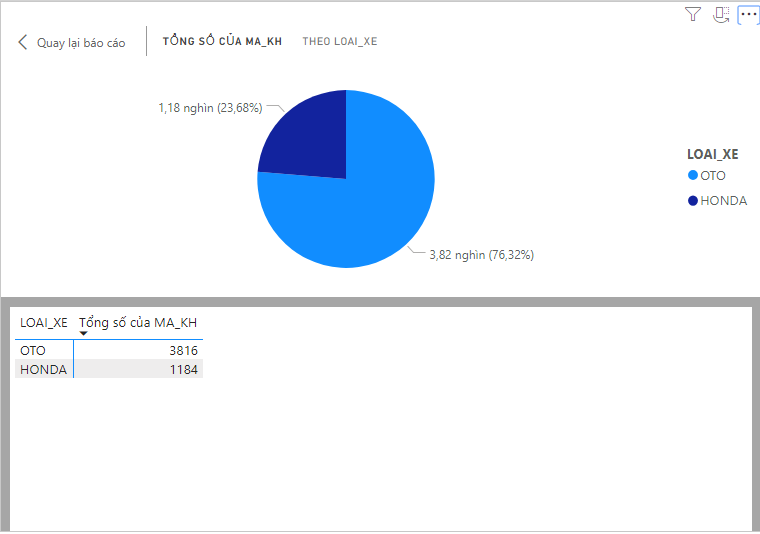


Hình . Thiết lập tham số mô hình luật kết hợp

# KẾT QUẢ PHÂN TÍCH

## Kết quả báo cáo BI

### Thống kê khách hàng theo loại xe



Hình . Thống kê khách hàng theo loại xe

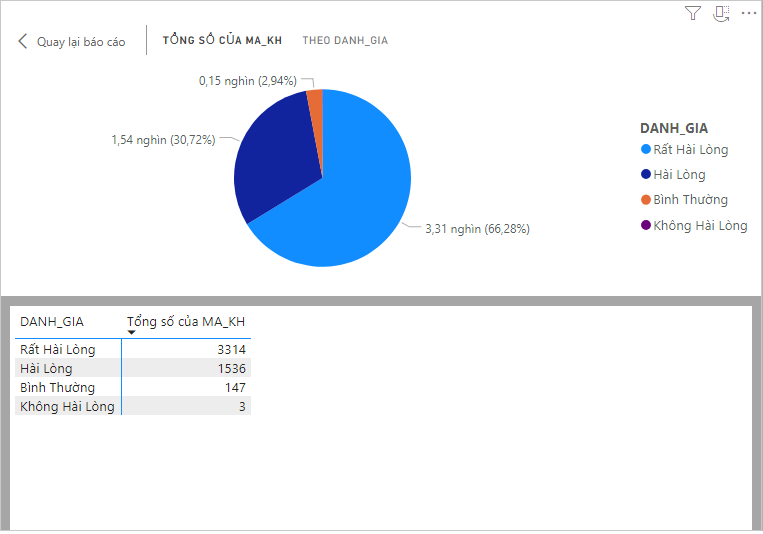
Nhìn vào biểu đồ ta thấy số tỉ lệ khách hàng sử dụng OTO vượt trội so với HONDA.

**Về loại xe**

OTO chiếm 76,32% với 3816 khách hàng.

HONDA chiếm 23,68% với 1184 khách hàng.

### Thống kê khách hàng theo đánh giá



Hình . Thống kê khách hàng theo đánh giá

Tỉ lệ khách hàng đánh giá Rất hài lòng chiếm đa số tiếp đó là Hài lòng và Bình thường. Tỉ lệ đánh giá Không hài lòng chiếm tỉ lệ rất nhỏ và không có khách hàng nào đánh giá Rất không hài lòng.

**Về đánh giá**

Rất hài lòng chiếm 66,28% với 3314 khách hàng.

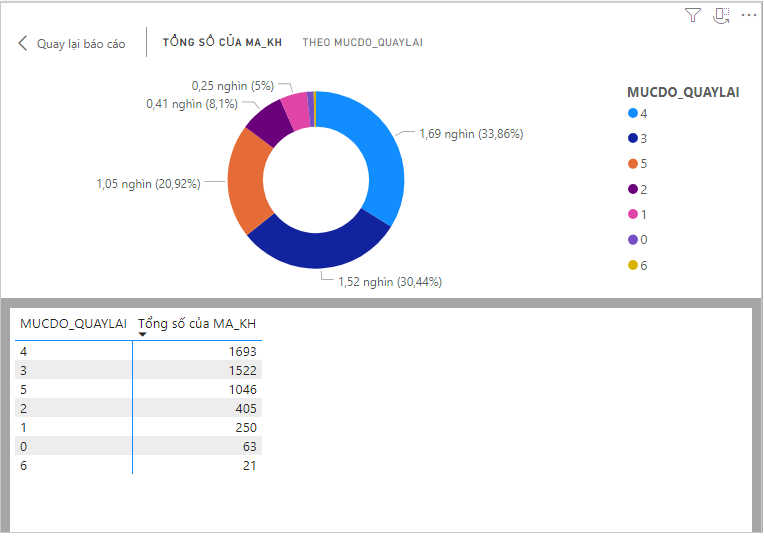
Hài lòng chiếm 30,72% với 1536 khách hàng.

Bình thường chiếm 2,94% với 147 khách hàng.

Không hài lòng chiếm 0,06% với 3 khách hàng.

Rất không hài lòng chiếm 0%.

### Thống kê khách hàng theo mức độ quay lại



Hình . Thống kê khách hàng theo mức độ quay lại

Dựa trên biểu đồ thì ta thấy mức độ quay lại của đa số khách hàng là 3 lần và 4 lần, tiếp đó là 5 lần và 2 lần ,1 lần, 0 lần và 6 lần chiếm tỉ lệ rất ít.

**Về mức độ quay lại**

4 lần chiếm 33,86% với 1692 khách hàng.

3 lần chiếm 30,44% với 1522 khách hàng.

5 lần chiếm 20,92% với 1046 khách hàng.

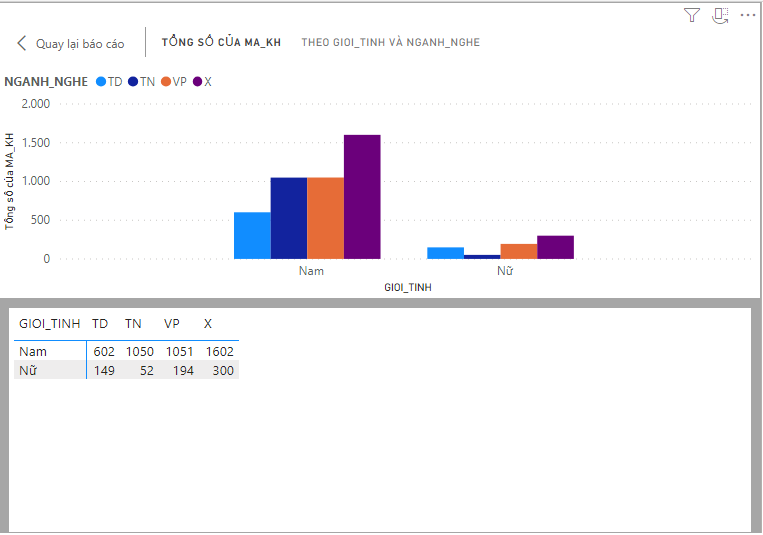
2 lần chiếm 8,1% với 405 khách hàng.

1 lần chiếm 5% với 250 khách hàng.

0 lần chiếm 1,26% với 63 khách hàng.

6 lần chiếm 0,42% với 21 khách hàng.

### Thống kê khách hàng theo giới tính và ngành nghề



Hình . Thống kê khách hàng theo giới tính và ngành nghề

**Về giới tính**

Nhìn chung khách hàng của chuỗi rữa xe là khách hàng nam, khách hàng nữ chiếm tỉ lệ tương đối thấp.

**Về ngành nghề**

Dựa vào biểu đồ trên thì ta thấy nhóm ngành X là sử dụng dịch vụ nhiều nhất và ở mỗi nhóm ngành thì khách hàng nam luôn chiếm đa số so với khách hàng nữ.

* **Chú thích**

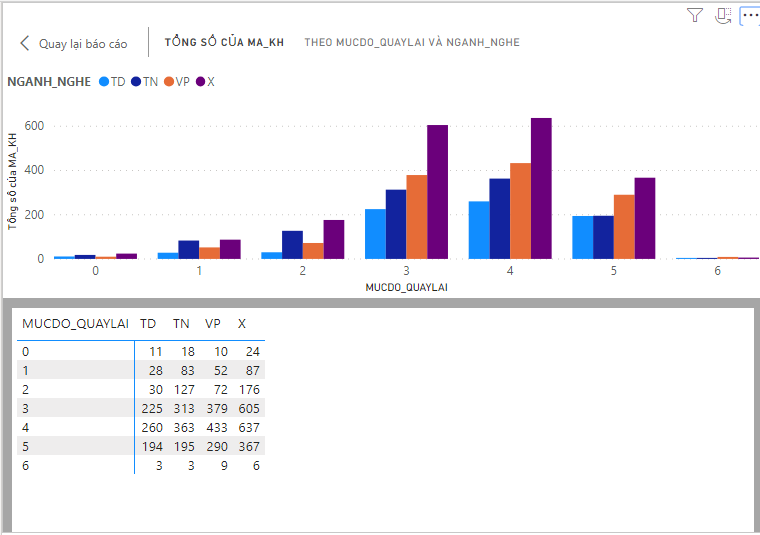
TD: Tự do

X: Không xác định

VP: Văn phòng

TN: Thương nhân

### Thống kê khách hàng theo mức độ quay lại và ngành nghề



Hình . Thống kê khách hàng theo mức độ quay lại và ngành nghề

**Về ngành nghề trong từng mức độ quay lại**

**0 lần:** mức độ quay lại 0 lần rất ít , trong đó ngành nghề X không quay lại có số lượng cao nhất với 24 khách hàng, tiếp đó là TN với 18 khách hàng, cuối cùng là VP và TD lần lượt là 10 và 11 khách hàng.

**1 lần:** tỉ lệ rất ít , trong đó nhóm ngành X và VP là cao nhất với lần lượt là 87 và 83 khách hàng, ngành nghề VP tương đối với 52 khách hàng, TD thấp nhất với 28 khách hàng.

**2 lần:** tương đối thấp trong các mức độ quay lại , ngành nghề X cao nhất với 176 khách hàng, tiếp đến là TN với 127 khách hàng, VP với 72 khách hàng và TD với 30 khách hàng.

**3 lần:** chiếm số lượng rất cao , trong đó ngành nghề X cao nhất với 605 khách hàng, tiếp đến là VP và TN lần lượt với 379 và 313 khách hàng, sau cùng là TD với 225 khách hàng.

**4 lần:** tỉ lệ rất cao , trong đó ngành nghề X chiếm cao nhất với 637 khách hàng, tiếp theo là VP và TN với 433 và 363 khách hàng, TD thấp nhất với 260 khách hàng.

**5 lần:** chiếm số lượng khá cao , trong đó ngành nghề X cao nhất với 367 khách hàng , VP đứng thứ 2 với 290 khách hàng, cuối cùng là TN và TD lần lượt với 195 và 194 khách hàng.

**6 lần**: chiếm tỉ lệ thấp nhất, trong đó ngành nghề X, VP, TN, TD có số lượng là 6, 9, 3, 3 khách hàng.

* **Chú thích**

Ngành nghề

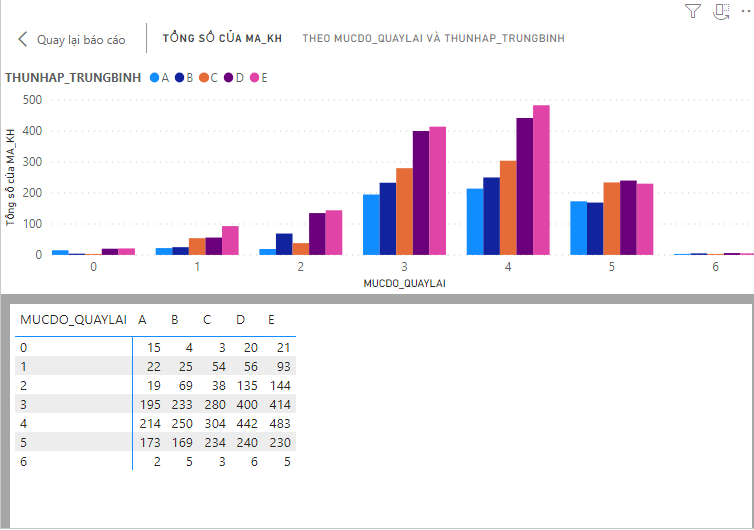
TD: Tự do

X: Không xác định

VP: Văn phòng

TN: Thương nhân

### Thống kê khách hàng theo mức độ quay lại và thu nhập trung bình



Hình . Thống kê khách hàng theo mức độ quay lại và thu nhập trung bình

**Về thu nhập trung bình trong từng mức độ quay lại**

**0 lần:** cao nhất là thu nhập trung trung bình E và D với lần lượt là 21 và 20 khách hàng, tiếp đến là A với 15 khách hàng ,cuối cung là B và C với 3 và 4 khách hàng.

**1 lần:** cao nhất là E với 93 khách hàng, D và C chiếm tương đối với 56 và 54 khách hàng, chiếm thấp nhất là A và B với 22 và 25 khách hàng.

**2 lần:** chiếm số lượng lớn nhất là E và D với 144 và 135 khách hàng, tiếp đến B với 69 khách hàng, C với 38 khách hàng, A với 19 khách hàng.

**3 lần:** cao nhất là E và D với 414 và 400 khách hàng , tiếp đến là C, B, A với lần lượt với 280, 233 và 195 khách hàng.

**4 lần:** cao nhất là E với 483 khách hàng, thứ hai là D với 442 khách hàng, tiếp theo là C với 304 khách hàng, cuối cùng là B và A với 250 và 214 khách hàng.

**5 lần:** C, D, E chiếm lần lượt là 234, 240, 230 khách hàng, tiếp theo là A và B với 173 và 169 khách hàng.

**6 lần:** số lượng khách hàng của thu nhập trung bình A, B, C, D, E lần lượt là 2, 5, 3, 6, 5 khách hàng.

* **Chú thích**

Thu nhập trung bình

A < 5tr (nhỏ hơn 5tr)

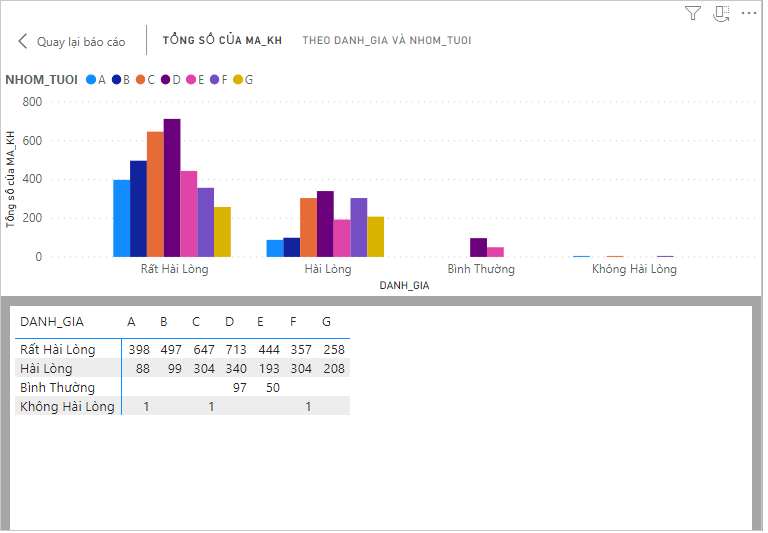
B: 5-10tr

C: 10 - 20tr

D: 20 - 30tr

E: > 30tr (lớn hơn 30tr)

### Thống kê khách hàng theo đánh giá và nhóm tuổi



Hình . Thống kê khách hàng theo đánh giá và nhóm tuổi

**Về nhóm tuổi trong từng đánh giá**

**Rất Hài Lòng:** Chiếm số lượng lớn nhất là nhóm tuổi D (713 khách hàng) và C (647 khách hàng). Chiêm số lượng nhỏ nhất là nhóm tuổi G (258 khách hàng) , các nhóm tuổi còn lại chiếm tương đối với A (398 khách hàng), B (497 khách hàng), E (444 khách hàng) , F (357 khách hàng).

**Hài Lòng:** Tương tự như rất hài lòng thì đánh giá hài lòng cũng chiếm cao nhất là nhóm tuổi D (340 khách hàng), tiếp đến là nhóm tuổi C (304 khách hàng) và F (304 khách hàng), chiếm cũng tương đối là nhóm tuổi G (208 khách hàng) và E (193 khách hàng), thấp nhất là nhóm tuổi B (99 khách hàng) và A (88 khách hàng).

**Bình Thường:** Chỉ có hai nhóm tuổi đánh giá là D (97 khách hàng), và E (50 khách hàng).

**Không Hài Lòng:** có ba nhóm tuổi đánh giá là A, C, F mỗi nhóm tuổi chỉ có 1 khách hàng.

**Rất Không Hài Lòng:** Không có nhóm tuổi nào đánh giá.

* **Chú thích**

Nhóm tuổi

A: < 20t (nhỏ hơn 20t)

B: 20 - 25t

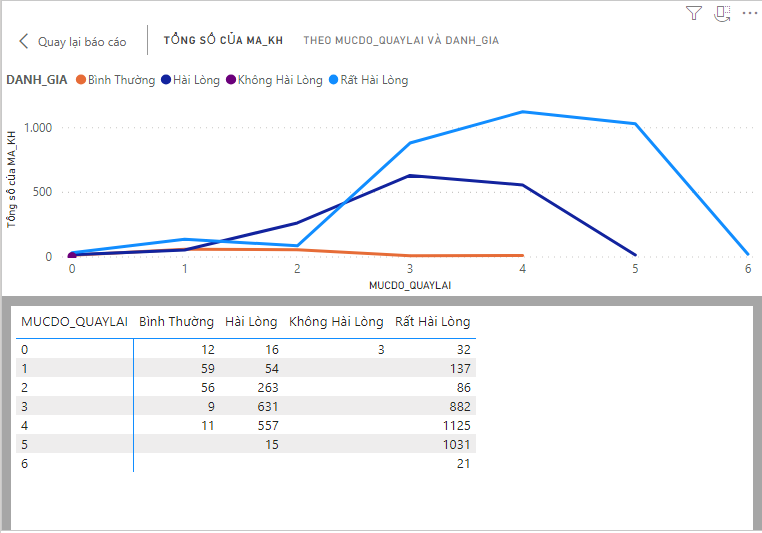
C: 25 - 30t

D: 30 - 35t

E: 35 - 40t

F: 40 - 45t

### Thống kê khách hàng theo mức độ quay lại và đánh giá



Hình . Thống kê khách hàng theo mức độ quay lại và đánh giá

**Về mức độ quay lại trong từng đánh giá**

**Rất Hài Lòng:** số khách hàng đánh giá rất hài lòng quay lại đa số từ 4 đến 5 lần với 1125 và 1031 khách hàng, mức độ quay lại với 3 lần cũng tương đối với 882 khách hàng.

**Hài Lòng:** số lượng đánh giá hài lòng quay lại sử dụng dịch vụ của VIETWASH từ 3 đến 4 lần với lần lượt là 631 và 557 khách hàng, số khách hàng quay lại 2 lần cũng khá tương đối với 263 khách hàng.

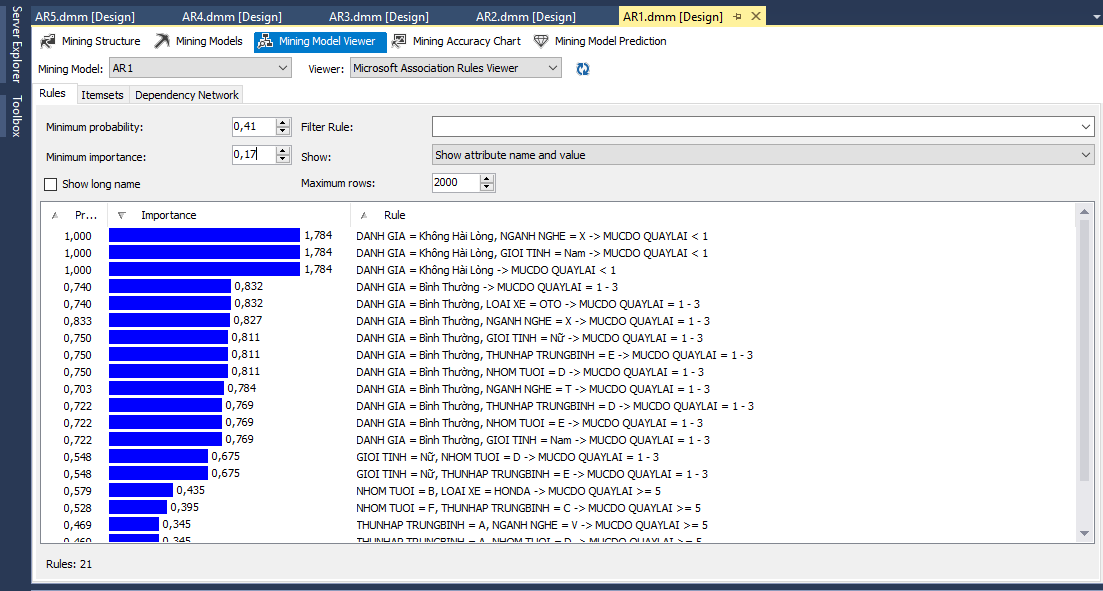
**Bình Thường:** số lượng quay lại sử dụng dịch vụ từ 1 đến 2 lần là chủ yếu với 59 và 56 khách hàng.

**Không Hài Lòng:** không có khách hàng nào quay lại sử dụng dịch vụ

**Rất Không Hài Lòng:** không có khách hàng nào đánh giá rất không hài lòng.

## Kết quả báo cáo luật kết hợp

### Kết quả mô hình AR1

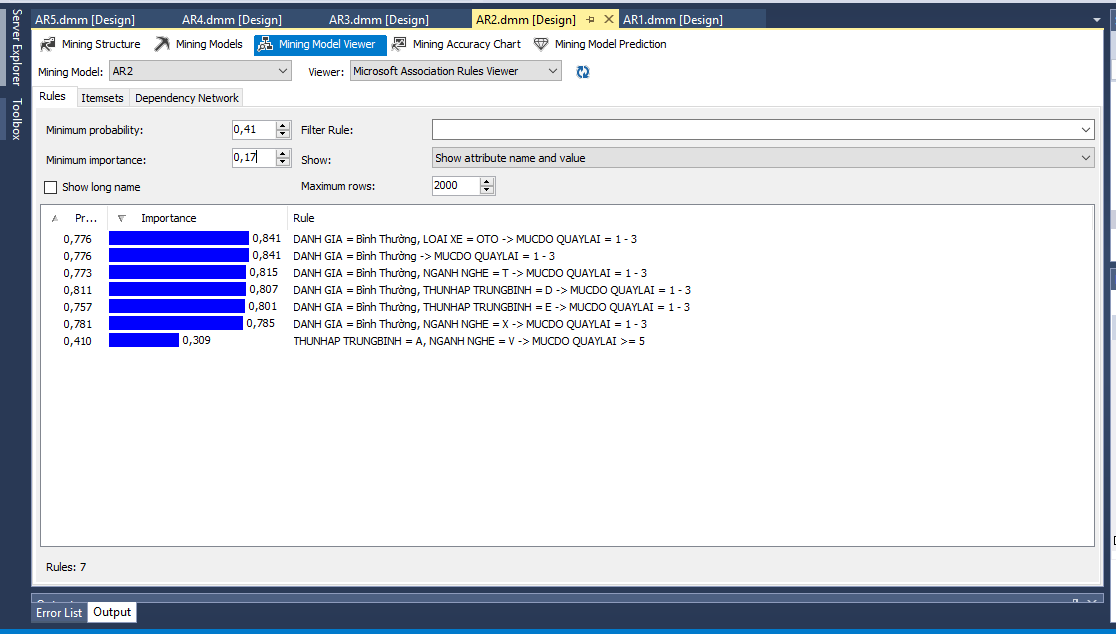


Hình . Kết quả mô hình AR1

***Các luật mạnh nhất:***

1. DANH GIA = Không Hài Lòng, NGANH NGHE = X -> MUCDO QUAYLAI < 1, xác suất 100%
2. DANH GIA = Không Hài Lòng, GIOI TINH = Nam -> MUCDO QUAYLAI < 1, xác suất 100%
3. DANH GIA = Không Hài Lòng -> MUCDO QUAYLAI < 1, xác suất 100%
4. DANH GIA = Bình Thường -> MUCDO QUAYLAI = 1 – 3, xác suất 74%
5. DANH GIA = Bình Thường, LOAI XE = OTO -> MUCDO QUAYLAI = 1 – 3, xác suất 74%
6. DANH GIA = Bình Thường, NGANH NGHE = X -> MUCDO QUAYLAI = 1 – 3, xác suất 83,3%
7. DANH GIA = Bình Thường, THUNHAP TRUNGBINH = E -> MUCDO QUAYLAI = 1 – 3, xác suất 75%
8. DANH GIA = Bình Thường, GIOI TINH = Nữ -> MUCDO QUAYLAI = 1 – 3, xác suất 75%
9. DANH GIA = Bình Thường, NHOM TUOI = D -> MUCDO QUAYLAI = 1 – 3, xác suất 75%
10. DANH GIA = Bình Thường, NGANH NGHE = TN -> MUCDO QUAYLAI = 1 – 3, xác suất 70,3%
11. DANH GIA = Bình Thường, THUNHAP TRUNGBINH = D -> MUCDO QUAYLAI = 1 – 3, xác suất 72,2%
12. DANH GIA = Bình Thường, GIOI TINH = Nam -> MUCDO QUAYLAI = 1 – 3, xác suất 72,2%
13. DANH GIA = Bình Thường, NHOM TUOI = E -> MUCDO QUAYLAI = 1 – 3, xác suất 72,2%
14. GIOI TINH = Nữ, NHOM TUOI = D -> MUCDO QUAYLAI = 1 – 3, xác suất 54,8%
15. GIOI TINH = Nữ, THUNHAP TRUNGBINH = E -> MUCDO QUAYLAI = 1 – 3, xác suất 54,8%
16. NHOM TUOI = B, LOAI XE = HONDA -> MUCDO QUAYLAI >= 5, xác suất 57,9%
17. NHOM TUOI = F, THUNHAP TRUNGBINH = C -> MUCDO QUAYLAI >= 5,xác suất 52,8%
18. THUNHAP TRUNGBINH = A, NHOM TUOI = D -> MUCDO QUAYLAI >= 5, xác suất 46,9%
19. THUNHAP TRUNGBINH = A, NGANH NGHE = VP -> MUCDO QUAYLAI >= 5, xác suất 46,9%
20. GIOI TINH = Nữ, THUNHAP TRUNGBINH = B -> MUCDO QUAYLAI >= 5, xác suất 43,1%
21. GIOI TINH = Nữ, LOAI XE = HONDA -> MUCDO QUAYLAI >= 5, xác suất 42,3%

### Kết quả mô hình AR2

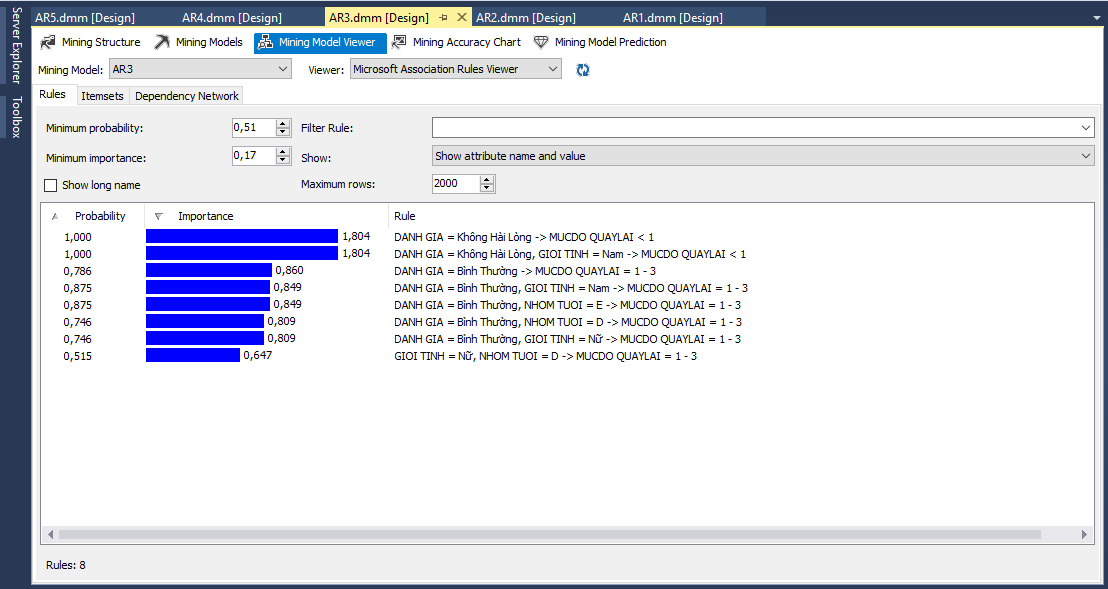


Hình . Kết quả mô hình AR2

***Các luật mạnh nhất:***

1. DANH GIA = Bình Thường -> MUCDO QUAYLAI = 1 – 3, xác suất 77,6%
2. DANH GIA = Bình Thường, LOAI XE = OTO -> MUCDO QUAYLAI = 1 – 3, xác suất 77,6%
3. DANH GIA = Bình Thường, NGANH NGHE = TN -> MUCDO QUAYLAI = 1 – 3, xác suất 77,3%
4. DANH GIA = Bình Thường, THUNHAP TRUNGBINH = D -> MUCDO QUAYLAI = 1 – 3, xác suất 81,1%
5. DANH GIA = Bình Thường, THUNHAP TRUNGBINH = E -> MUCDO QUAYLAI = 1 – 3, xác suất 75,7%
6. DANH GIA = Bình Thường, NGANH NGHE = X -> MUCDO QUAYLAI = 1 – 3, xác suất 78,1%
7. THUNHAP TRUNGBINH = A, NGANH NGHE = VP -> MUCDO QUAYLAI >= 5, xác suất 41%

### Kết quả mô hình AR3

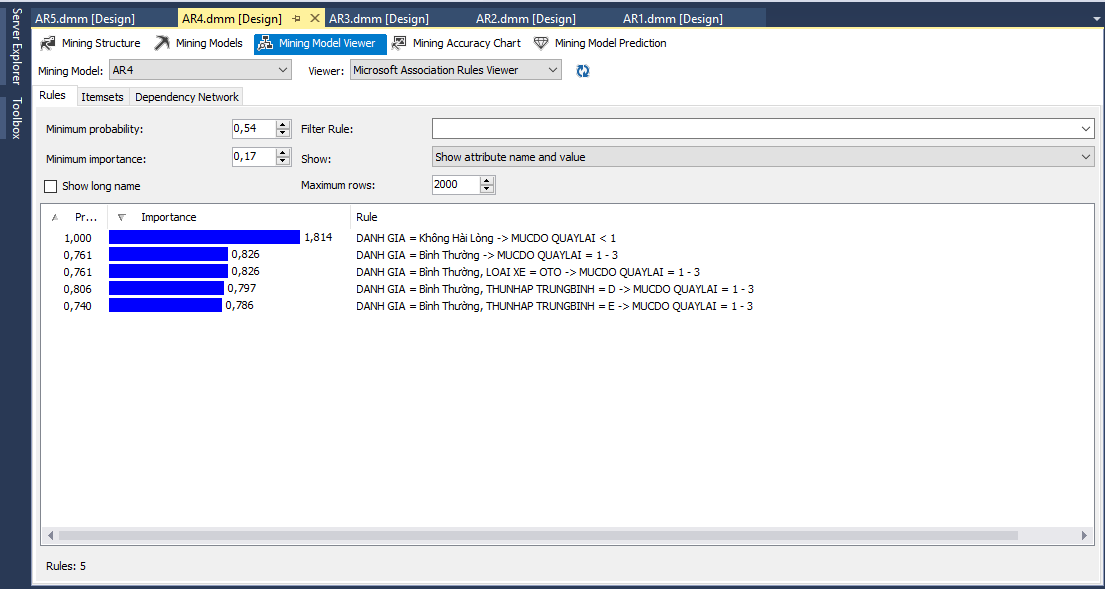


Hình . Kết quả mô hình AR3

***Các luật kết hợp mạnh nhất:***

1. DANH GIA = Không Hài Lòng, GIOI TINH = Nam -> MUCDO QUAYLAI < 1, xác suất 100%
2. DANH GIA = Không Hài Lòng -> MUCDO QUAYLAI < 1, xác suất 100%
3. DANH GIA = Bình Thường -> MUCDO QUAYLAI = 1 – 3, xác suất 78,6%
4. DANH GIA = Bình Thường, GIOI TINH = Nam -> MUCDO QUAYLAI = 1 – 3, xác suất 87,5%
5. DANH GIA = Bình Thường, THUNHAP TRUNGBINH = D -> MUCDO QUAYLAI = 1 – 3, xác suất 87,5%
6. DANH GIA = Bình Thường, GIOI TINH = Nữ -> MUCDO QUAYLAI = 1 – 3, xác suất 74,6%
7. DANH GIA = Bình Thường, THUNHAP TRUNGBINH = E -> MUCDO QUAYLAI = 1 – 3, xác suất 74,6%
8. GIOI TINH = Nữ, THUNHAP TRUNGBINH = E -> MUCDO QUAYLAI = 1 – 3, xác suất 51,5%

### Kết quả mô hình AR4

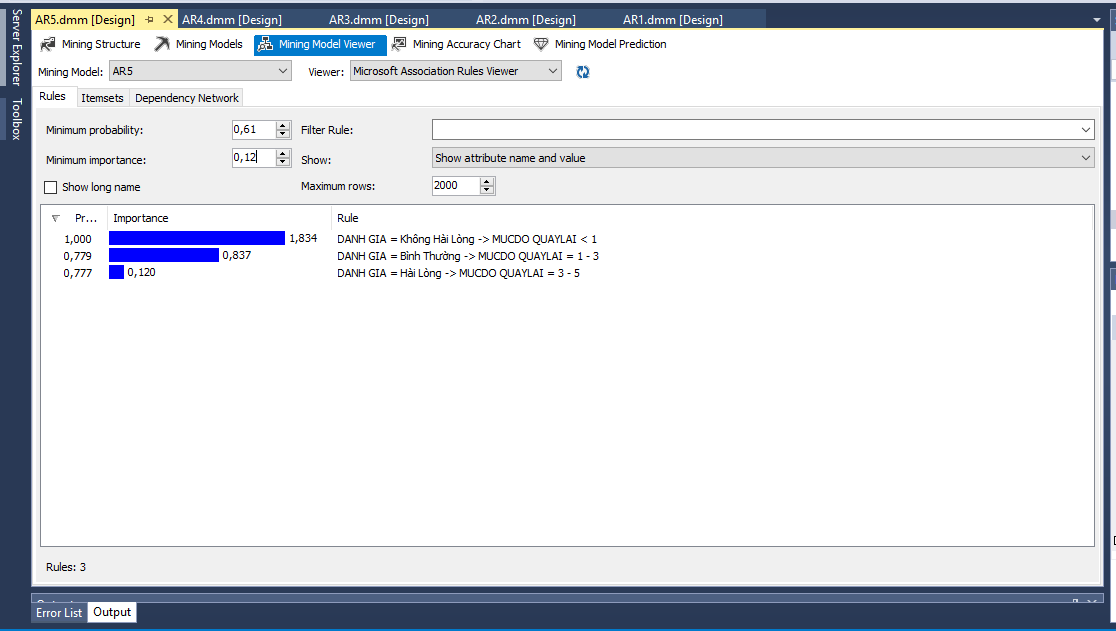


Hình 5.12 Kết quả mô hình AR4

Các luật kết hợp mạnh nhất

1. DANH GIA = Không Hài Lòng -> MUCDO QUAYLAI < 1, xác suất 100%
2. DANH GIA = Bình Thường -> MUCDO QUAYLAI = 1 – 3, xác suất 100%
3. DANH GIA = Bình Thường, LOAI XE = OTO -> MUCDO QUAYLAI = 1 – 3, xác suất 76,1%
4. DANH GIA = Bình Thường, THUNHAP TRUNGBINH = D -> MUCDO QUAYLAI = 1 – 3, xác suất 80,6%
5. DANH GIA = Bình Thường, THUNHAP TRUNGBINH = E -> MUCDO QUAYLAI = 1 – 3, xác suất 74%

### Kết quả mô hình AR5



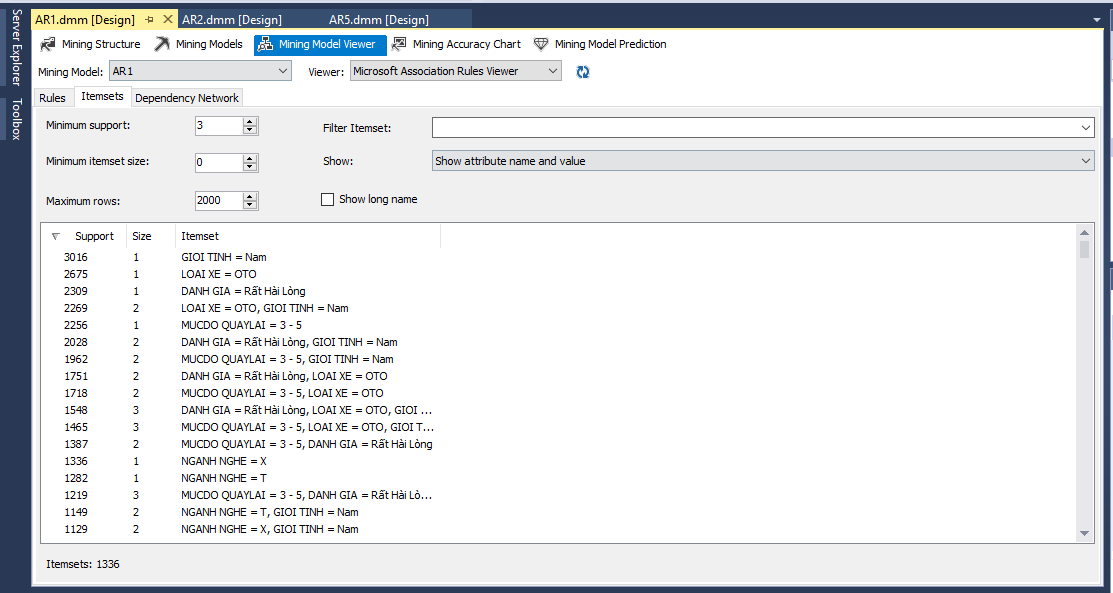
Hình . Kết quả mô hình AR5

***Các luật kết hợp mạnh nhất:***

1. DANH GIA = Không Hài Lòng -> MUCDO QUAYLAI < 1, xác suất 100%
2. DANH GIA = Bình Thường -> MUCDO QUAYLAI = 1 – 3, xác suất 77,9%
3. DANH GIA = Hài Lòng -> MUCDO QUAYLAI = 3 – 5, xác suất 77,7%

### Itemsets

Itemsets cho biết các thông tin quan trọng của luật kết hợp như Support (độ hỗ  
trợ của luật kết hợp), Size (Số items trong Itemsets):



Hình . Itemsets của mô hình

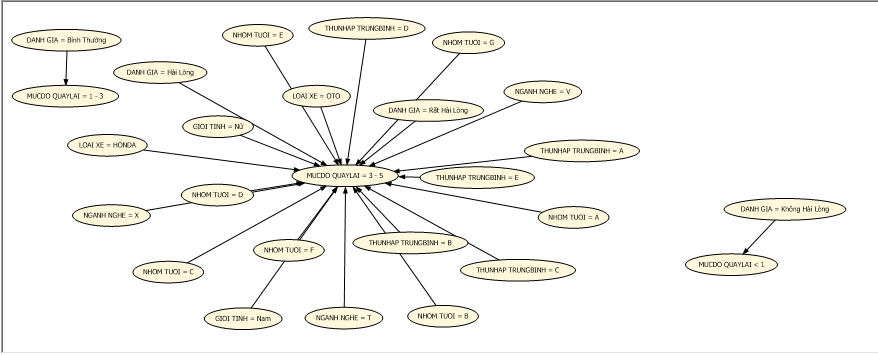
Bảng Itemset chỉ ra rõ được độ phổ biến của các tập hợp thông qua độ hỗ trợ, độ hỗ trợ càng cao thì mức độ phổ biến của tập hợp càng lớn, cũng như biết được số lượng các thuộc tính trong mỗi tập thông qua size. Từ đó có cái nhìn tổng quan hơn về kết quả nghiên cứu.

### Mạng phụ thuộc (Dependency Network)

Sử dụng Dependency Network cho phép hiểu được sự tác động của các items  
khác nhau trong Model. Mỗi Node trong Dependency Net thể hiện một Item, bằng  
cách chọn một item, giao diện sẽ hiện lên các items khác được xác định bởi Item đã  
chọn (hoặc dùng để xác định Item đã chọn) trong model.

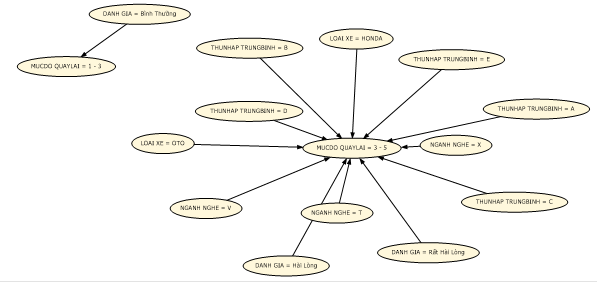
Có thể kéo thanh trượt bên trái để xem các mức độ kết hợp (mạnh hay yếu) giữ  
các Items trong model.

Mô hình AR1



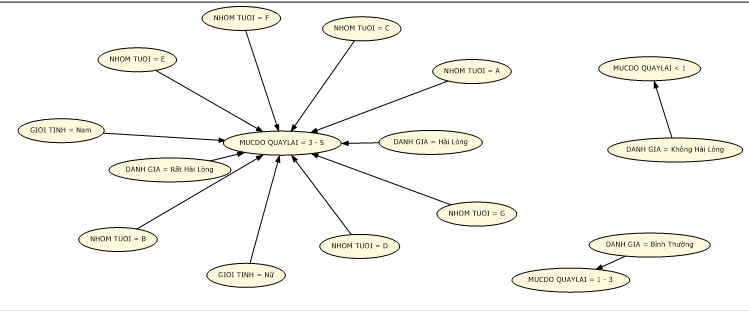
Hình . Mạng phụ thuộc (Mô hình AR1)

Mô hình AR2



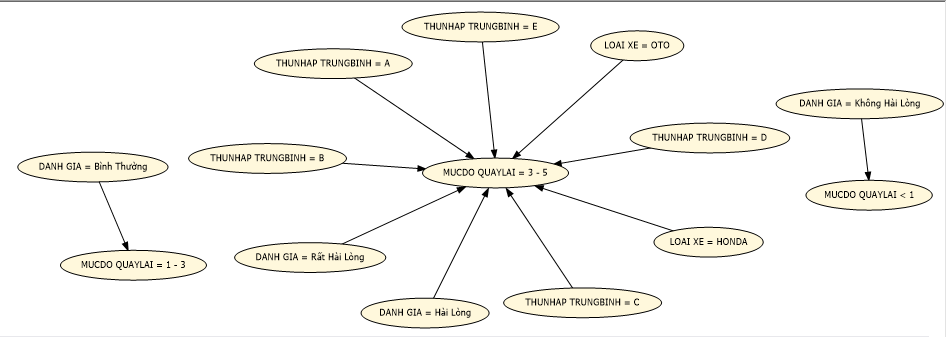
Hình . Mạng phụ thuộc (Mô hình AR2)

Mô hình AR3



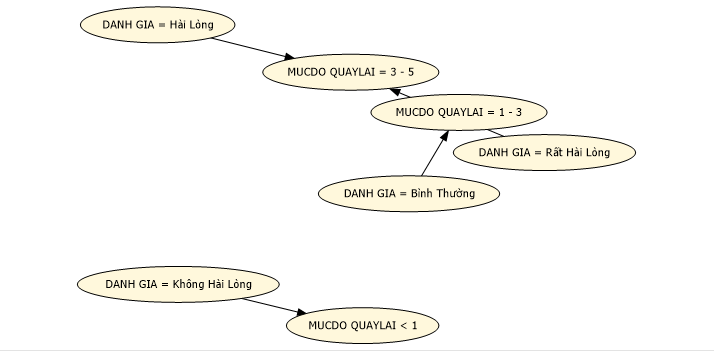
Hình . Mạng phụ thuộc (Mô hình AR3)

Mô hình AR4



Hình . Mạng phụ thuộc (Mô hình AR4)

Mô hình AR5



Hình . Mạng phụ thuộc (Mô hình AR5)

## Tổng kết phân tích

Theo kết quả phân tích BI thì đa số khách hàng của công ty là nam, độ tuổi đa dạng, ngành nghề chủ yếu là thương nhân và nhân viên văn phòng, loại xe sử dụng là OTO, mức độ đánh giá đa số là hài lòng, mức độ quay lại là từ 3 đến 5 lần là chủ yếu.

Theo kết quả báo cáo luật kết hợp tổng thể ta thấy đánh giá Hài lòng và Rất hài lòng thì mức độ quay lại 3 đến 5 lần, đánh giá bình thường thì mức độ quay lại từ 1 đến 3 lần, đánh giá không hài lòng thì mức độ quay lại dưới 1 lần.

Nhìn chung dịch vụ rửa xe theo chuỗi là một dịch vụ mới nên đa số phản hồi của khách hàng là tích cực, tuy nhiên cũng cần có nhiều chính sách marketing để thu hút nhiều đối tượng khách hàng hơn ngoài khối văn phòng và thương nhân, bên cạnh đó loại xe hay đến sử dụng dịch vụ là OTO tỉ lệ chiếm tới hơn 76% trong khi HONDA chỉ chiếm khoảng 26%, điều này liên quan đến chính sách giá còn chưa hợp lý và hình ảnh chưa thân thiện với đa số khách hàng bình dân. Cuối cùng cần thực hiện chính sách chăm sóc khách hàng hợp lý để gia tăng được mức độ hài lòng và mức độ quay lại của khách hàng, tung ra nhiều ưu đãi và khuyến mãi đối với khách hàng trung thành.

KẾT LUẬN

Bài báo cáo đã đưa ra được những cơ sở lý thuyết của Bussiness Intelligent, Khai phá dữ liệu và Luật kết hợp đồng thời nghiên cứu về các công cụ như Power BI, Microsoft SQL Server 2014, Microsoft SQL Server Analysis Services (SSAS), Môi trường triển khai ứng dụng (BIDS). Bằng những cơ sơ lý thuyết trên cùng với dữ liệu do EFOX Solution cung cấp, bài báo cáo cũng đã phân tích cụ thể các số liệu thu thập được từ cuộc khảo sát, qua đó có thể tìm hiểu về mức độ hài lòng của khách hàng đối với sản phẩm dịch vụ , giúp doanh nghiệp có những chính sách, kế hoạch phù hợp hơn nhằm nâng cao trải nghiệm cũng như đánh giá của khách hàng về sau.

Bên cạnh những mặt đạt được được thì có những khía cạnh khác còn chưa tốt như:

Các luật đưa ra chưa đa dạng còn bị trùng nhau, phần lớn luật có độ mạnh không cao nên bị lược bỏ nhiều nguyên nhân do số liệu khảo sát khá khiêm tốn (khoảng 5000 dòng).

TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | N. M. Tiến, "VIBLO," 26 09 2016. [Online]. Available: https://viblo.asia/p/tong-quan-ve-business-intelligent-bi-DZrGNNXwGVB. |
| [2] | Đ. T. D. Ngọc. [Online]. Available: http://uet.vnu.edu.vn/~thuyhq/Student\_Thesis/K44\_Do\_Thi\_Dieu\_Ngoc\_Thesis.pdf. |
| [3] | V. T. Minh, TÌM HIỂU LUẬT KẾT HỢP, Ha Noi: 2010, 2010. |
| [4] | "BAC," 22 08 2019. [Online]. Available: http://www.bacs.vn/vi/blog/cong-cu-ho-tro/power-bi-la-gi-4919.html. |
| [5] | "timoday," [Online]. Available: https://timoday.edu.vn/bai-1-tong-quan-ve-sql-server/. |

PHỤ LỤC