**LỜI CAM ĐOAN**

Em xin cam đoan chuyên đề tốt nghiệp **“*Nghiên cứu tableau và ứng dụng cho bài toán phân khúc khách hàng*”** là chuyên đề nghiên cứu độc lập dưới sự hướng dẫn của cô Nguyễn Thị Hương Lý. Các nội dung trong để tài này được em thu thập từ các nguồn cung cấp dữ liệu cho nghiên cứu và học tập. Tất cả mọi thứ liên quan đến nội dung, cũng như kết quả thu được trong để tài này là trung thực và chưa được công bố ở bất kỳ đâu.

Nếu như có bất kỳ sự gian lận nào em xin hoàn toàn chịu trách nhiệm trước bộ môn, hội đồng, khoa Công nghệ thông tin về nội dung báo cáo của mình, cũng như các hình thức kỷ luật theo quy định mà Nhà trường đã đề ra.

*Khánh Hòa, ngày tháng năm 2024*

**Sinh viên cam kết**

**Nguyễn Sơn Tùng**

**LỜI CẢM ƠN**

Lời đầu tiên em xin chân thành cảm ơn tất cả giảng viên đến từ Khoa Công nghệ thông tin, những người thầy, người cô là người thầm lặng suốt những năm tháng, đã chỉ bảo và cho em những bài học vô giá và kiến thức rất bổ ích và cũng cảm ơn đến các bạn lớp 62.TTQL-2.

Và quan trọng hơn em xin gửi lời cảm ơn chân thành đến giảng viên ThS. Nguyễn Thị Hương Lý, là người cô đã hướng dẫn và tạo mọi điều kiện giúp đỡ, chỉ ra những cái sai của em trong suốt ba tháng làm chuyên đề tốt nghiệp vừa qua.

Em xin cảm ơn đến gia đình, bạn bè của em đã động viên, cổ vũ và đóng góp nhiều ý kiến trong quá trình học tập, đời sống cá nhân cũng như quá trình thực hiện chuyên đề. Do thời gian còn hạn chế và kiến thức bản thân của em còn nhiều hạn chế nên chuyên đề tốt nghiệp của em thực hiện sẽ không tránh khỏi những sai lầm thiếu sót. Rất mong nhận được những góp ý từ thầy cô giảng viên và các bạn để em có thêm nhiều kinh nghiệm trong phân tích và thiết kế để chuẩn bị thật tốt cho công việc sau này.

Em xin chân thành cảm ơn!

Mục Lục

**CHƯƠNG I: CƠ SỞ LÝ THUYẾT**

**1.1. Tổng quan về BI**

Business Intelligence (BI): là giải pháp quản trị doanh nghiệp thông minh, là một hệ thống báo cáo cho phép các tổ chức, doanh nghiệp khai thác dữ liệu từ nhiều nguồn khác nhau như về khách hàng, thị trường, nhà cung cấp, đối tác, nhân sự, … phân tích và sử dụng các dữ liệu đó thành các nguồn thông tin có ý nghĩa nhằm hỗ trợ việc ra quyết định.

Hệ thống BI gồm 6 thành phần chính (hình 1.1):

- Data Sources (nguồn dữ liệu): 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, phần mềm bán hàng, website thương mại điện tử, …

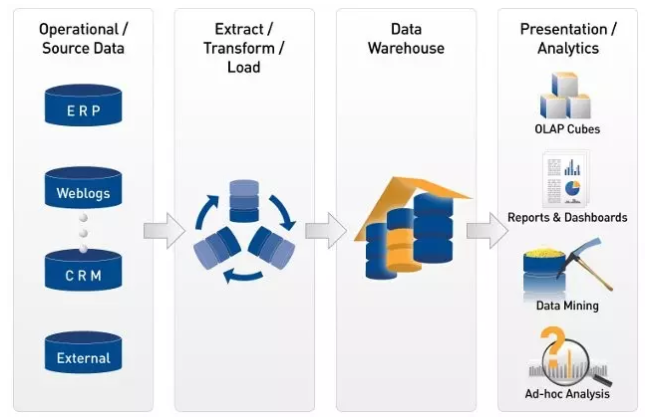
- Data Warehouse (kho dữ liệu): là cơ sở dữ liệu tích hợp, chứa dữ liệu tổng hợp của doanh nghiệp.

- Data Exploration (khám phá dữ liệu): là công việc đầu tiên khi chúng ta bắt đầu vào phân tích, tìm hiểu dữ liệu.

- Data Mining (khai phá dữ liệu): các kỹ thuật dùng để khai phá dữ liệu và phát hiện tri thức như phân loại, phân cụm, phát hiện luật kết hợp, dự báo, …

- Data Presentations & Visualization (trực quan hoá dữ liệu): tạo ra các báo cáo, biểu đồ từ quá trình data mining để phục vụ cho nhu cầu của người dùng cuối,…

- Decisions Making (ra quyết định): là quá trình bao gồm nhiều bước liên quan đến việc xác định vấn đề, hình thành, lựa chọn phương án thông qua văn bản quyết định.



Hình 1. . Các thành phần của 1 hệ thống BI

1.3. Dashboard

Dashboard thường được lập ra bởi các nhà phân tích kinh tế chuyên nghiệp. Trong một dashboard có rất nhiều báo cáo (Report), trong đó (Dashboard là bảng thông tin tổng hợp kết hợp nhiều báo cáo (Report) trên một màn hình hiển thị, giúp cho người xem có cái nhìn tổng quát và nhanh chóng có nhiều thông tin quản trị khác nhau.

**Báo cáo (Report)** đơn giản là những báo cáo quản trị thể hiện các số liệu ở nhiều nội dung và nhiều góc độ khác nhau tuỳ theo nhu cầu của người quản lý.

Graphical user interface, application, Word

Description automatically generated

Hình 1. 2. Báo cáo doanh thu của sản phẩm

**Vậy Dashboard là gì?** Dashboard có thể hiểu được là một bảng điều khiển kỹ thuật số (digital control), hay còn được coi là một giao diện số được dùng để thu thập và tổng hợp dữ liệu của toà bộ tổ chức.

“Trong dashboard không chỉ cung cấp các dữ liệu chuyên sâu trong quá trình sản xuất kinh doanh, đồng thời dashboard còn đưa ra một cái nhìn tổng quát về năng suất cuẩ từng bộ phận, các xu hướng, các hoạt động, các chỉ số **KPI (Key Performance Indicator** – hay còn gọi là chỉ số đánh giá thực hiện công việc).”

Graphical user interface, application

Description automatically generated

Hình 1. 3. Tổng hợp doanh thu trên Dashboard

Như vậy, Dashboard đưa ra cách nhìn tổng quan về mọi mặt trong hoạt động kinh doanh của tổ chức sao cho việc trình bày phải dễ hiểu và toàn diện chỉ trong 1 trang báo cáo.

Nhà quản lý có thể thấy được các thước đo chỉ số hoạt động của doanh nghiệp mình bằng hình ảnh, đồ thị đa chiều.

Một số ưu điểm mà Dashboard mang lại:

- Trực quan và sinh động giúp cho các nhà phân tích có thể tìm ra vấn đề một cách nhanh chóng.

- Giảm áp lực cho người trình bày khi đọc vì báo cáo chỉ tóm gọn trong một màn hình trình chiếu hoặc một trang giấy.

- Linh hoạt, dễ dàng cho phép người dùng tương tác để lựa chọn,

- Trình bày thông tin mang tính hỗ trợ đưa ra quyết định.

- Tiết kiệm thời gian lập và trình bày báo cáo vì tính tự động hoá của Dashboard qua việc tạo và lập trên những form mẫu có sẵn trên máy tính.

**1.4. Tiền xử lý dữ liệu**

Chúng ta thường bắt đầu việc xử lý dữ liệu thô, chứ không có dữ liệu chuẩn bị sẵn. Phần lớn công việc xây dựng một kho dữ liệu là trích chọn, làm sạch và chuyển đổi dữ liệu.

Dữ liệu có chất lượng cao nếu như phù hợp với mục đích sử dụng trong điều hành, ra quyết định và lập kế hoạch.

Các công việc của tiền xử lý dữ liệu:

+ Làm sạch dữ liệu (data cleaning): loại bỏ nhiễu, hiệu chỉnh những phần dữ liệu không nhất quán.

+ Tích hợp dữ liệu (data integration): trộn dữ liệu nhiều nguồn khác nhau vào một kho dữ liệu.

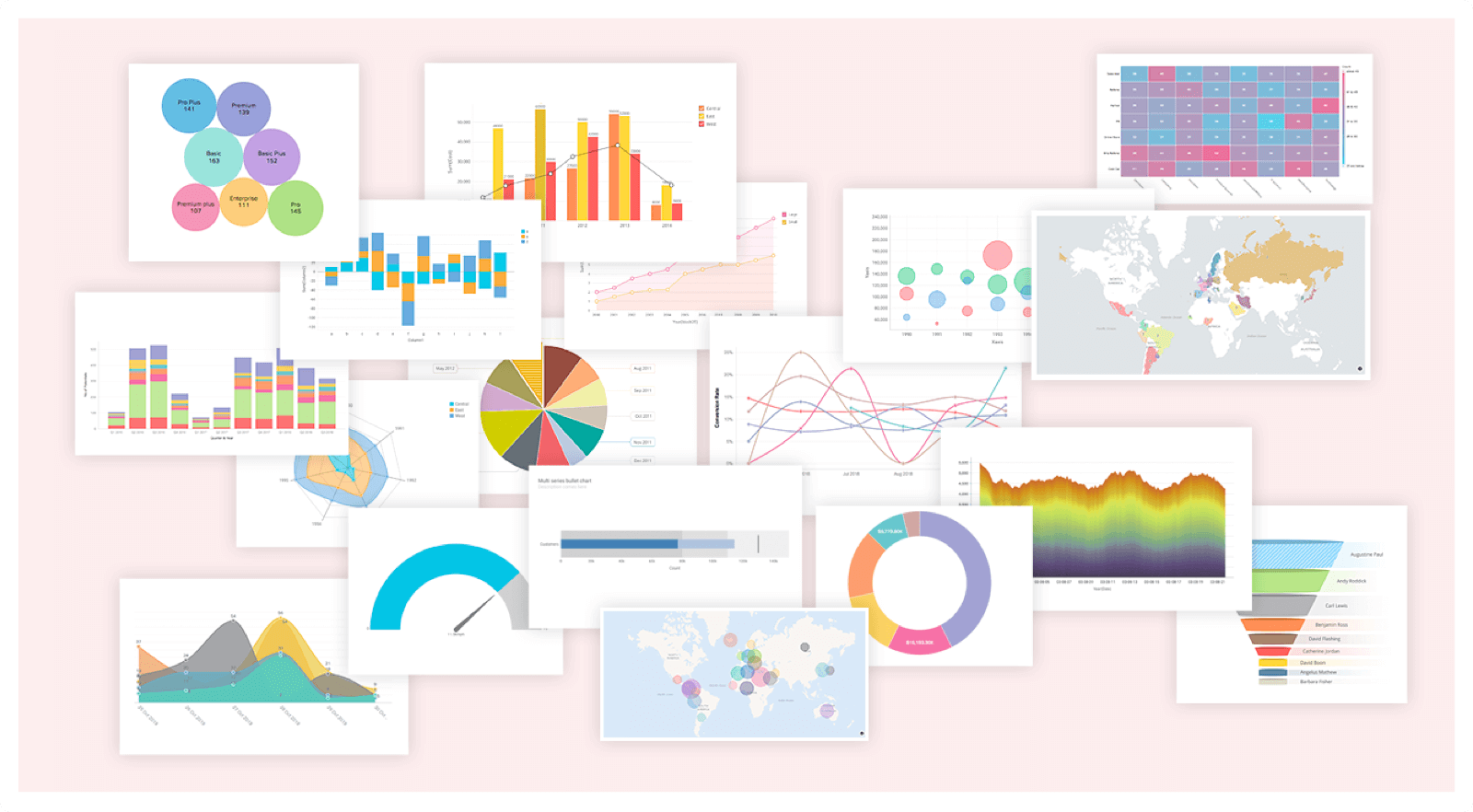
+ Chuyển đổi dữ liệu (data transformation): là chuẩn hoá dữ liệu (data normalization).

+ Rút gọn dữ liệu (data reduction): thu giảm kích thước dữ liệu (giảm số phần tử) bằng cách kết hợp dữ liệu (data aggregtion), loại bỏ các đặc điểm dư thừa (giảm số chiều/ thuộc tính dữ liệu), gom cụm dữ liệu.

**1.5. Trực quan hoá dữ liệu**

Là một phương thức biểu diễn trực quan của dữ liệu dưới các hình ảnh, biểu đồ, bảng đồ và đồ thị trực quan để truyền tải thông tin đến người dùng một cách dễ hiểu hơn, sinh động hơn.

Sfdfs sdf sdfsdf sđfsd sé s Sfdfs sdf sdfsdf sđfsd sé s sdfsdf sđfsd sé s Sfdfs sdf sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s sdfsdf sđfsd sé s



*Hình 1. 4. Các biểu đồ trực quan hoá dữ liệu*

**1.6. Phân khúc khách hàng sử dụng RFM**

“Từ những khách hàng tiêu dùng trung thành cho đến những khách hàng rời bỏ công ty, tất cả những khách hàng đều có nhu cầu và mong muốn đa dạng. Công ty muốn khách hàng chi tiêu nhiều hơn từ những chiến dịch tiếp thị chương trình, sản phẩm mới tới khách hàng theo những cách khác nhau. Tuy nhiên, câu hỏi đặt ra là làm thế nào để đưa ra được các chiến dịch tiếp thị phù hợp với những nhóm khách hàng đang có nhu cầu để từ đó tăng tỷ lệ phản hồi từ khách hàng và từ đó tăng doanh số bán hàng. Bài toán đặt ra là làm thế nào để có thể phân khúc khách hàng một cách tương đối chính xác dựa trên hành vi giao dịch lịch sử của khách hàng, thuật toán RFM sẽ giúp chúng ta giải quyết vấn đề này một cách nhanh chóng và hiệu quả”.

**1.6.1 Tìm hiểu RFM**

Mô hình này được gọi là RFM, là một kỹ thuật tiếp thị được sử dụng để đánh giá hành vi của khách hàng của công ty. Ví dụ: nó có thể được sử dụng để đánh giá thời gian gần đây khách hàng đã mua sản phẩm hoặc dịch vụ (lần truy cập gần đây), số lần khách hàng đã mua hàng kể từ một ngày (tần suất) nhất định và khách hàng đã chi bao nhiêu trong một ngày nhất định một khoảng thời gian (tiền tệ).

- Lần mua hàng gần đây – R (Recency).

- Tần suất mua hàng của khách hàng – F (Frequency).

- Tiền, giá trị, khách hàng chi bao nhiêu tiền – M (Monetary value).

RFM là một mô hình rất phổ biến vì các lý do:

- Nó sử dụng thang đo một cách khách quan.

- Doanh nghiệp sử dụng mô hình này một cách hiệu quả và dễ dàng vì đây là mô hình đơn giản.

- Trực quan - đầu ra của phương pháp phân khúc này là dễ hiểu và giải thích.

**1.6.2. Lợi ích của RFM**

*a. Hiểu về khách hàng của bạn*

Nhờ vào mô hình RFM, các công ty hiểu hơn về khách hàng của họ: phân loại được nhóm khách hàng trung thành của họ, những khách hàng không hài lòng với dịch vụ của họ, khách hàng cũ đã rời bỏ họ và từ đó hiểu hơn về sản phẩm của họ.

*b. Tăng doanh số của bạn*

“Bằng cách xem xét khách hàng thông qua ba khía cạnh chính của mô hình phân tích RFM, doanh nghiệp của chúng ta có thể cải thiện đáng kể các kỹ thuật tiếp thị của mình và nhắm mục tiêu đưa ra các chiến dịch của doanh nghiệp mình một cách chính xác hơn theo mong muốn và nhu cầu của khách hàng, điều này cuối cùng sẽ tăng tỷ lệ phản hồi và doanh số bán hàng”.

*c. Giảm đáng kể lượng khách hàng rời bỏ công ty/ doanh nghiệp*

Với phân tích RFM, doanh nghiệp có thể phân loại khách hàng của mình thành một số danh mục. Số lượng danh mục phụ thuộc vào bản chất của doanh nghiệp và bản năng của nhà phân tích. Từ kết quả phân tích doanh nghiệp có thể tìm ra nhóm khách hàng trung thành cũng như nhóm khách hàng có khả năng rời bỏ công ty hoặc đã rời bỏ công ty và để lấy lại được lượng khách hàng rời bỏ công ty thì công ty hay doanh nghiệp cần có các hành động khác nhau để lấy lại khách hàng. Ví dụ: Công ty hoặc doanh nghiệp có thể gửi e-mail cá nhân cho khách hàng đến những người có nguy cơ rời đi để kết nối lại với doanh nghiệp.

*d. Quy tắc 80/20*

Đây là quy tắc ngón tay cái tuyên bố rằng 80% kết quả có thể được quy cho 20% nguyên nhân cho một sự kiện nhất định. Thông thường, một doanh nghiệp sử dụng phương pháp này để chỉ ra rằng 80% doanh thu được tạo ra bởi 20% tổng số khách hàng của mình. Cách tiếp cận này giúp những người ra quyết định tập trung vào các yếu tố quan trọng nhất ảnh hưởng đến tài nguyên của họ. Ví dụ (hình 1.5):

Chart

Description automatically generated

*Hình 1. 5. Biểu đồ minh hoạ nguyên tắc Pareto (nguyên tắc 80/20).*

Nếu doanh nghiệp có thể xác định những người chi tiêu hàng đầu của mình, doanh nghiệp có thể đưa ra các chiến lược tiếp thị phù hợp để giữ chân khách hàng hiện tại hoặc có được những khách hàng có hồ sơ tương tự.

Tuy nhiên so với RFM thì quy tắc 80/20 vẫn còn hạn chế vì nó chỉ nhìn vào tần suất chi tiêu, và không giải quyết được sự suy thoái.

**CHƯƠNG II: TỔNG QUAN VỀ CÔNG CỤ BI TABLEAU**

**2.1. Tableau BI**

**2.1.1. Giới thiệu về Tableau**

Tableau là một phần mền được tạo bởi một công ty cùng tên ở Washington và được thành lập vào năm 2003. Tableau là công cụ phần mềm hỗ trợ phân tích và trực quan hoá dữ liệu (Data Visualization), được dùng nhiều trong lĩnh vực BI (Business Intelligence). Cũng giống như Excel, Tableau giúp tổng hợp các dữ liệu nhưng ở một cấp độ cao hơn khi chuyển những dữ liệu này từ các dãy số văn bản thành những hình ảnh, biểu đồ trực quan xây dựng các dashboard và phân tích.

Tableau thiết kế cho tất cả mọi người sử dụng một cách dễ dàng, không yêu cầu mọi người phải giỏi về lập trình. Vì vậy, bất cứ một ai cũng có thể sử dụng Tableau cho những vấn đề họ cần tìm hiểu về các doanh nghiệp mà họ muốn với các thao tác kéo và thả chuột linh hoạt. Giao diện Tableau dễ dàng sử dụng mà bất cứ người dùng nào thành thạo Excel thì có thể tìm hiểu và sử dụng các thao tác này bằng chuột một cách linh hoạt.

**2.1.2. Các tính năng của Tableau**

Khi sử dụng Tableau người dùng có thể thực hiện một số các chức năng :

- Mô phỏng và phân tích dữ liệu.

- Phân tích dữ liệu với Big Data.

- Phân tích theo thời gian.

- Hỗ trợ tạo các truy vấn bằng thao tác đơn giản.

- Nhập dữ liệu với kích thước lớn, quản lý siêu dữ liệu.

- Kéo thả để xây dựng các Business Dashboard từ đơn giản đến phức tạp.

- Kết nối và chia sẻ thông qua các ứng dụng trực tuyến thời gian thực.

- Biên dịch các truy vấn thành những hình ảnh, biểu đồ.

- Quản lý các dữ liệu hoạt động (Data Stories).

**2.1.3. Lợi ích khi sử dụng Tableau**

Sử dụng BI Tableau sẽ có các lợi ích :

- Tổng hợp và phân tích dữ liệu từ nhiều nguồn, khả năng tích hợp dữ liệu đa dạng và đưa ra các góc nhìn về hoạt động doanh nghiệp.

- Các nhà doanh nghiệp, quản trị luôn có những chuỗi dữ liệu thời gian giúp họ nhận ra các sự khác biệt và các phương án ứng xử đó. Đồng thời giúp họ đưa ra nền tảng và quyết định chính xác và đúng thời điểm.

- Tableau xây dựng một môi trường làm việc dựa trên nền tảng số liệu, trên nền tảng phân tích, tạo nên một sực mạnh tập thể cho doanh nghiệp.

- Ai cũng có thể xem và phân tích và các số liệu của mình quản lý, làm cho doanh nghiệp của người dùng vận hành trong môi trường ưu tú nhất và hiệu quả nhất.

- Tableau được triển khai nhanh, khả năng tiếp cận nhanh, giảm tối đa chi phí và dễ dàng mở rộng theo một mô hình.

**2.1.4. Các công cụ của Tableau**

Tableau hỗ trợ nhiều công cụ cho người dùng lựa chọn:

**- Tableau Prep :** là công cụ được dùng để chuẩn bị dữ liệu, trích xuất và xử lý dữ liệu (ETL). Mang đến sự thay đổi quan trọng trong việc tổ chức dữ liệu, so với phương pháp truyền thống có nhiều cải tiến. Tableau Prep giúp mọi người dùng doanh nghiệp và nhà phân tích định hình dữ liệu nhanh chóng, cho phép thực hiện các truy vấn, kết hợp và làm sạch dữ liệu cực kì đơn giản và tiện lợi. Sử dụng Tableau Prep giúp dữ liệu có tổ chức, rõ ràng, dễ quản lý hơn. Hiện nay có hai công cụ là Tableau Prep Builder để xây dựng luồng dữ liệu và Tableau Prep Conductor để quản lý các luồng.

**- Tableau Desktop** cung cấp giao diện trực quan cùng các tính năng đa dạng mã hoá và phân tích dữ liệu. Phần lớn là các thao tác kéo thả chuột và cũng không yêu cầu quá nhiều việc về mặt kỹ thuật hay lập trình. Tableau Desktop có khả năng kết nối rộng rãi đến nhiều định dạng file khác nhau, để đáp ứng tốt nhất về nhu cầu phân tích trong nhiều ngành nghề, lĩnh vực.

**- Tableau Online :** là nền tảng phân tích của bạn được lưu trữ hoàn toàn trên đám mây. Xuất bản trang tổng quan và chia sẻ khám phá của bạn với bất kỳ ai. Mời đồng nghiệp hoặc khách hàng tìm hiểu và khám phá các cơ hội tiểm ẩn bằng hình ảnh tương tác và dữ liệu chính xác. Tất cả đều có thể truy cập dễ dàng từ trình duyệt hoặc khi đang di chuyển bằng các ứng dụng dành cho điện thoại di động. Phần lớn sử dụng Tableau online chủ yếu sử dụng trong việc học tập.

**- Tableau Server :** mở rộng giá trị dữ liệu trên toàn bộ tổ chức với máy chủ tableau. Là nơi chia sẻ các báo cáo phân tích của doanh nghiệp, những dữ liệu trên đây được bảo mật cẩn thận và cấp quyền truy cập.