

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA CÔNG NGHỆ THÔNG TIN



# BÁO CÁO ĐỒ ÁN TRỰC QUAN HÓA DỮ LIỆU

| TLAB03 |

ỨNG DỤNG VÀ CÔNG CỤ  
TRỰC QUAN HÓA DỮ LIỆU

| GIẢNG VIÊN HƯỚNG DẪN |  
THẦY LÊ NHỰT NAM

| NHÓM 1 |

21120242 – Huỳnh Thị Kiều Hoa  
21120247 – Nguyễn Văn Quang Hưng  
21120254 – Lưu Chấn Huy  
21120298 – Chiêm Bình Nguyên  
21120302 – Huỳnh Trí Nhân

TP. Hồ Chí Minh - 2024

# MỤC LỤC

MỤC LỤC .....	2
1. THÔNG TIN CHUNG .....	3
1.1 Check-list đồ án.....	3
1.2 Bảng phân công công việc.....	3
2. TÌM HIỂU VỀ CÔNG CỤ TABLEAU .....	4
2.1 Giới thiệu về Tableau .....	4
2.2 Tính năng của Tableau .....	4
2.3 Ưu, khuyết điểm .....	9
3. Phụ lục.....	11
TÀI LIỆU THAM KHẢO .....	12

# 1. THÔNG TIN CHUNG

## 1.1 Check-list đồ án

- ☒ Tìm hiểu về công cụ
- ☒ Trực quan hóa trên tập dữ liệu
- ☒ Rút ra ý nghĩa sau khi trực quan
- ☒ Sử dụng mô hình học máy cơ bản
- ☒ Báo cáo
- ☒ Video

## 1.2 Bảng phân công công việc

Tên	Phân công công việc	Mức độ hoàn thành	Mức độ đóng góp
Kiều Hoa	Tìm hiểu công cụ, báo cáo, thiết kế khung cho phần mềm	100%	100%
Quang Hưng	Thực hiện trực quan hóa với Tableau, quay video demo.	100%	100%
Chấn Huy	Thực hiện trực quan hóa với Tableau, báo cáo tính năng công cụ	100%	100%
Bình Nguyên	Thực hiện trực quan hóa với Tableau, báo cáo tính năng công cụ	100%	100%
Trí Nhân	Thực hiện trực quan hóa với Tableau, nhận xét ưu khuyết điểm	100%	100%

## 2. TÌM HIỂU VỀ CÔNG CỤ TABLEAU

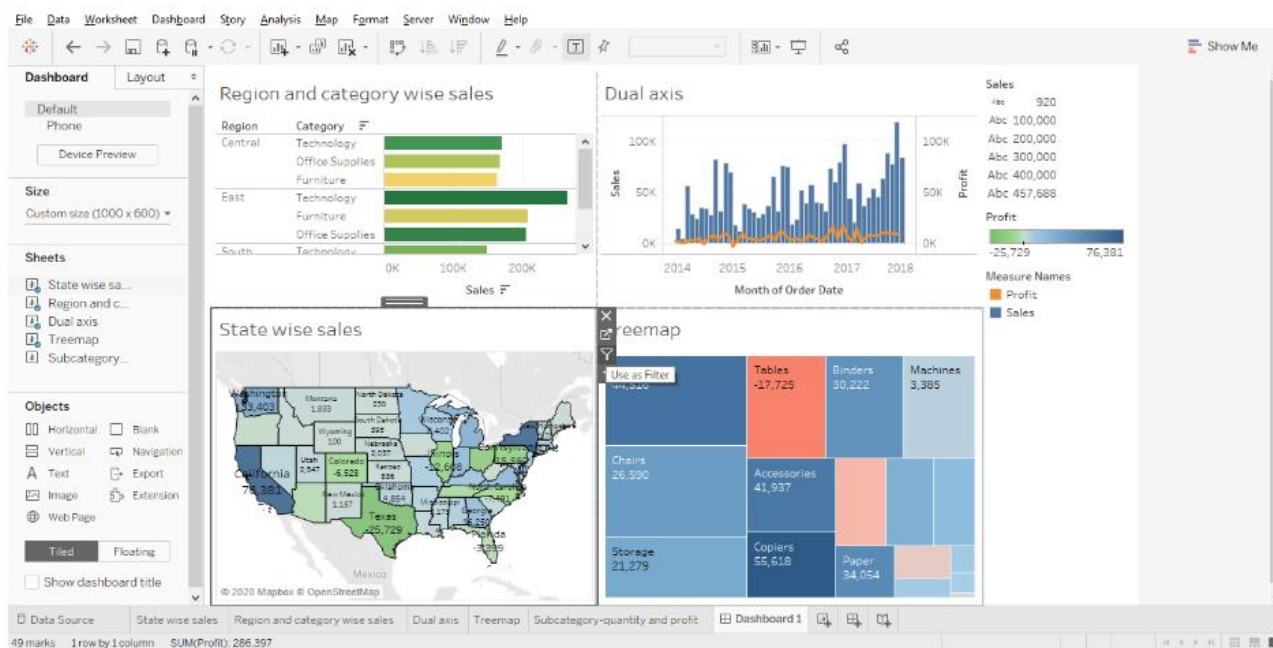
### 2.1 Giới thiệu về Tableau

- Tableau là một công cụ phần mềm hỗ trợ phân tích và trực quan hóa dữ liệu, giúp kết nối, tổng hợp dữ liệu từ các nguồn dưới dạng các biểu đồ một cách trực quan nhất cho người dùng. Sở hữu nhiều tính năng độc đáo và thú vị, Tableau được thiết kế để giúp người dùng dễ dàng hiểu và trình bày dữ liệu theo một cách hấp dẫn. Phần mềm này tận dụng các lĩnh vực như học máy, thống kê, ngôn ngữ tự nhiên và chuẩn bị dữ liệu thông minh để hỗ trợ và gia tăng khả năng sáng tạo của con người trong quá trình phân tích.
- Tableau được công nhận là TOP 1 trong Gartner Magic Quadrant\* về BI Tool bảy năm liên tiếp từ 2012 đến 2019. Tableau mang đến đầy đủ ứng dụng phục vụ nhu cầu phân tích & trực quan hóa dữ liệu.

### 2.2 Tính năng của Tableau

#### 2.2.1. Trực quan hóa dữ liệu

##### 2.2.1.1. Biểu đồ và đồ thị



- Tableau là một công cụ trực quan hóa dữ liệu vô cùng mạnh mẽ, cung cấp chúng ta có khả năng vẽ hàng loạt biểu đồ khác nhau phù hợp với dữ liệu từ những biểu đồ đơn giản như: biểu đồ cột, đường, tròn, scatter, gantt,... Cho đến những biểu đồ có khả năng phân tích dữ liệu nâng cao: heatmapp, bản đồ, historam, waterfall chart, sankey diagram,...

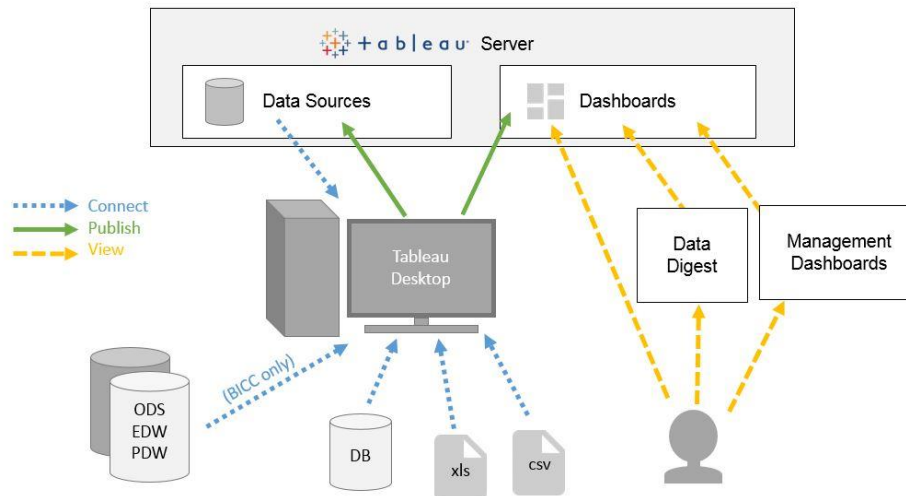
- **Biểu đồ và đồ thị** giúp chúng ta biến dữ liệu số và văn bản thành hình ảnh, giúp người xem nhanh chóng hiểu được các xu hướng, mẫu, và các điểm bất thường trong dữ liệu cùng với nhiều lợi ích đi kèm:
  - **Dễ hiểu hơn** : Với hàng ngàn dòng, hàng triệu dòng chỉ cô đọng trong một vài biểu đồ giúp chúng ta nắm bắt thông tin hiệu quả hơn, đổi thông tin phức tạp thành hình ảnh đơn giản, dễ hiểu. Điều này đặc biệt hữu ích trong việc trình bày dữ liệu cho những người không quen thuộc với số liệu chi tiết.
  - **Tăng tính tương tác**: Tableau cho phép tạo ra các biểu đồ và đồ thị tương tác, giúp người dùng khám phá dữ liệu bằng cách tương tác với các visualizations trực tiếp. **Để dàng đưa ra quyết định dựa trên dữ liệu**: Với trực quan hóa dữ liệu, doanh nghiệp có thể nhanh chóng phát hiện các xu hướng, so sánh hiệu suất, và xác định các cơ hội kinh doanh, giúp ra quyết định nhanh chóng và chính xác.
  - **Kể chuyện với dữ liệu**: Biểu đồ và đồ thị trong Tableau cho phép chúng ta kể một câu chuyện dữ liệu, giúp người dùng hiểu rõ hơn về bối cảnh và động lực của dữ liệu.
  - **Phân tích đa chiều**: Tableau hỗ trợ nhiều loại biểu đồ và đồ thị, cho phép chúng ta phân tích dữ liệu theo nhiều chiều khác nhau, từ đó mang lại cái nhìn toàn diện về tình hình kinh doanh.

#### 2.2.1.2.Dashboards

- Bằng các kết hợp nhiều biểu đồ chúng ta tạo nên những **Dashboard**. **Dashboard** trong Tableau là một bảng điều khiển của Tableau bao gồm một tập hợp các báo cáo, thành phần tương tác và graphics đồ họa được sắp xếp và trình bày trên cùng một màn hình. Nó cho phép người dùng kết hợp dữ liệu từ nhiều nguồn khác nhau để tạo ra một cái nhìn tổng quan về các số liệu, xu hướng và insights quan trọng. Từ đó mang đến nhiều lợi ích cho doanh nghiệp và những nhà phân tích dữ liệu như:
  - **Storytelling**: Chúng ta có thể kể một câu chuyện bằng cách liên kết các visualizations trên dashboard, trình bày dữ liệu theo trình tự logic để giúp những người hiểu rõ hơn về bối cảnh và những gì chúng ta muốn truyền đạt
  - **Dashboard tương tác được**: Tableau cho phép người dùng tương tác với các visualizations trên dashboard. Ví dụ: chúng ta có thể nhấp vào một phần của biểu đồ để lọc dữ liệu trong phần còn lại của dashboard hoặc để đi sâu hơn vào chi tiết cụ thể.
  - **Lọc dữ liệu**: chúng ta có thể thêm các bộ lọc để người dùng có thể nhanh chóng điều chỉnh dữ liệu họ muốn xem, từ đó giúp họ phân tích dữ liệu theo nhiều cách khác nhau.
  - **Đào sâu vào dữ liệu**: Cho phép người dùng đi sâu vào chi tiết dữ liệu bằng cách nhấp vào các visualizations và mở ra các bảng biểu hoặc biểu đồ mới với thông tin chi tiết hơn.
  - **Tích hợp dữ liệu**: Tableau cho phép chúng ta kết nối với nhiều nguồn dữ liệu khác nhau và hiển thị chúng trong cùng một dashboard, giúp chúng ta có một bức tranh toàn cảnh về dữ liệu của mình.
  - **Tính linh hoạt và tùy chỉnh**: Người dùng có thể sắp xếp và customize các thành phần trên bảng điều khiển của Tableau theo cách phù hợp nhất với nhu cầu của họ.

#### 2.2.2. Kết nối dữ liệu

### 2.2.2.1. Kết nối database

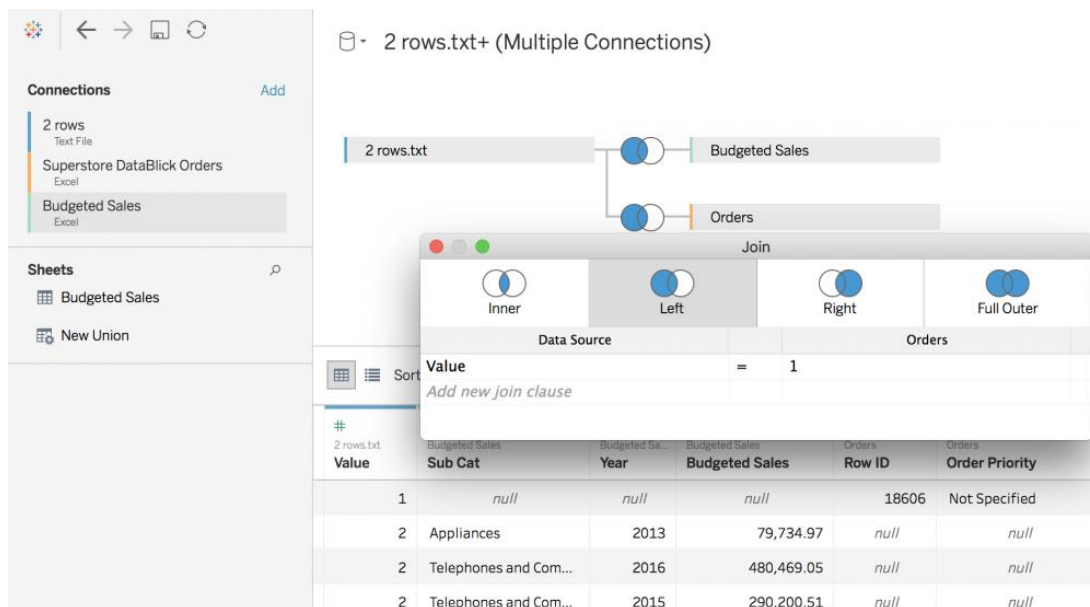


- Trong Tableau hỗ trợ kết nối rất nhiều kiểu dữ liệu khác nhau như là : Excel files, JSON, CSV files, SQL Databases, ... Không chỉ một tập dữ liệu đơn thuần mà Tableau hỗ trợ rất nhiều tệp dữ liệu được nạp vào cùng lúc, chúng ta có thể kết nối và xử lý những tệp dữ liệu đó trên chính Tableau của chúng ta

### 2.2.2.2. Kết nối dữ liệu Cloud

- Tableau cung cấp khả năng kết nối với nhiều nguồn dữ liệu đám mây khác nhau, giúp người dùng dễ dàng phân tích và trực quan hóa dữ liệu mà không cần tải về máy:
  - **Hỗ trợ đa dạng nguồn dữ liệu đám mây:** Tableau hỗ trợ nhiều nguồn dữ liệu đám mây phổ biến như Amazon Redshift, Google BigQuery, Snowflake, Microsoft Azure SQL Data Warehouse, và các dịch vụ khác. Điều này cho phép chúng ta truy cập vào nhiều loại dữ liệu đám mây để phân tích.
  - **Tự động làm mới dữ liệu:** Tableau có khả năng tự động làm mới dữ liệu từ các nguồn đám mây. Điều này giúp chúng ta luôn có dữ liệu mới nhất để phân tích, mà không cần cập nhật thủ công.
  - **Bảo mật:** Tableau cung cấp các tùy chọn bảo mật tích hợp để bảo vệ dữ liệu khi kết nối với các nguồn đám mây. chúng ta có thể kiểm soát quyền truy cập, sử dụng mã hóa, và các tính năng khác để bảo vệ dữ liệu của mình.
  - **Hiệu suất:** Tableau tối ưu hóa hiệu suất truy vấn dữ liệu từ các nguồn đám mây bằng cách sử dụng các phương pháp như extract (trích xuất dữ liệu), live connection (kết nối trực tiếp), và in-memory analytics (phân tích trong bộ nhớ). Điều này giúp tối ưu hóa hiệu suất truy vấn ngay cả khi dữ liệu rất lớn.
  - **Khả năng mở rộng:** Khả năng kết nối dữ liệu đám mây của Tableau giúp doanh nghiệp mở rộng quy mô phân tích dữ liệu một cách dễ dàng. Chúng ta có thể truy cập vào dữ liệu từ bất kỳ đâu và từ bất kỳ thiết bị nào.
  - **Tích hợp với các dịch vụ khác:** Tableau cũng tích hợp tốt với các dịch vụ đám mây khác như Salesforce, Google Analytics, và các nền tảng đám mây khác, giúp chúng ta kết hợp dữ liệu từ nhiều nguồn khác nhau.

### 2.2.2.3. Kết hợp dữ liệu



Phương pháp	Mô tả
Relationship	<ul style="list-style-type: none"> <li>Thiết lập tiềm năng để nối giữa hai bảng dựa trên các trường liên quan. Không hợp nhất dữ liệu lại để tạo thành một bảng cố định mới.</li> <li>Trong quá trình phân tích, Tableau sẽ tự động truy vấn các bảng liên quan bằng cách sử dụng các liên kết phù hợp theo ngữ cảnh để tạo ra một bảng dữ liệu tùy chỉnh cho phân tích đó.</li> <li>Giữ nguyên mức độ chi tiết phù hợp, không làm mất dữ liệu, giữ nguyên các phép tổng hợp phù hợp và xử lý các giá trị null.</li> </ul>
Joins	<ul style="list-style-type: none"> <li>Kết hợp hai bảng dữ liệu dựa trên điều kiện liên kết và loại liên kết để tạo thành một bảng dữ liệu cố định mới.</li> <li>Thường được sử dụng để thêm các cột dữ liệu mới trong cùng một cấu trúc hàng cơ bản.</li> <li>Có thể gây mất dữ liệu với một số loại liên kết nếu các trường không có trong tất cả các bảng.</li> <li>Có thể gây trùng lặp dữ liệu nếu các bảng có mức độ chi tiết khác nhau.</li> </ul>
Unions	<ul style="list-style-type: none"> <li>Kết hợp hai hoặc nhiều bảng dữ liệu để tạo thành một bảng dữ liệu cố định mới.</li> <li>Được sử dụng để thêm các hàng dữ liệu mới vào cấu trúc cột cơ bản giống nhau.</li> </ul>
Blends	<ul style="list-style-type: none"> <li>Làm việc trên hai hoặc nhiều nguồn dữ liệu Tableau riêng biệt.</li> <li>Dữ liệu vẫn giữ nguyên.</li> <li>Tableau truy vấn độc lập các nguồn dữ liệu và trực quan hóa kết quả cùng nhau trong view, dựa trên các trường liên kết được thiết lập cho sheet đó.</li> </ul>

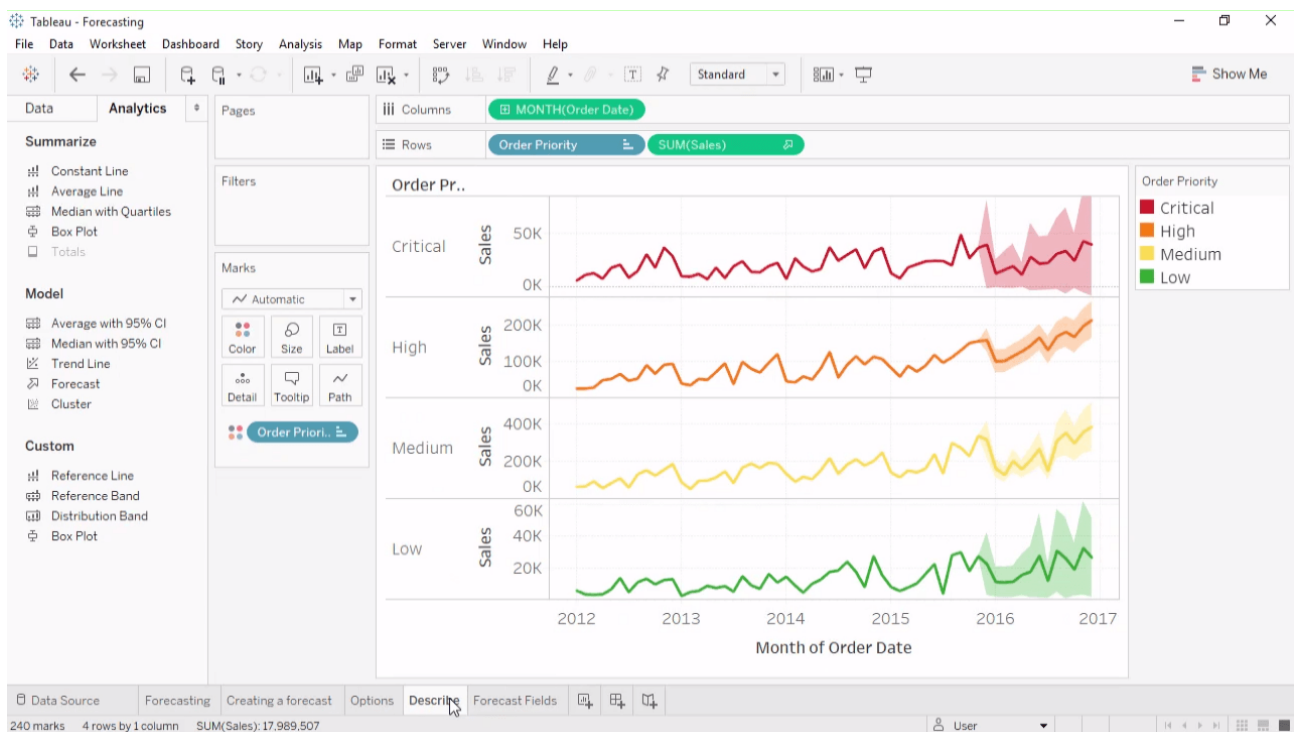


## 2.2.3. Phân tích dữ liệu

### 2.2.3.1. Khả năng tính toán

- Tableau cung cấp khả năng tính toán vô cùng mạnh mẽ, đây là một công cụ mạnh mẽ cho phép người dùng tạo ra các trường mới từ dữ liệu hiện có thông qua các công thức toán học, logic, hoặc các hàm tích hợp sẵn. Điều này cho phép người dùng phân tích và hiển thị dữ liệu theo các cách thức phù hợp với nhu cầu cụ thể của họ
- Có rất nhiều hàm được Tableau cung cấp để hỗ trợ người dùng tính toán trên từng loại dữ liệu: String function, date function, logic function, user function,...

### 2.2.3.2. Dự đoán



- Dự đoán **Time Series**: Tableau sử dụng các mô hình thống kê để dự báo xu hướng trong chuỗi thời gian, như dự đoán doanh thu trong tương lai dựa trên dữ liệu lịch sử.

### 2.2.3.3. Xây dựng mô hình thống kê

- **Hồi quy**: Phân tích mối quan hệ giữa các biến số, tìm hiểu mối quan hệ giữa dữ liệu độc lập và phụ thuộc. Liệu chúng có quan hệ tuyến tính với nhau hay không tức là từ một giá ta có thể dự đoán ra đại lượng còn lại.
- **Phân cụm**: Tableau có tính năng K-Means clustering để tìm hiểu các cụm tự nhiên trong dữ liệu, giúp chúng ta phân nhóm dữ liệu theo các đặc trưng tương đồng.

## 2.2.4. Khả năng chia sẻ

- Phân tích dữ liệu không nên là một nhiệm vụ cô lập - đó là lý do tại sao Tableau được xây dựng để hỗ trợ sự cộng tác. Các thành viên trong nhóm có thể chia sẻ dữ liệu, thực hiện các

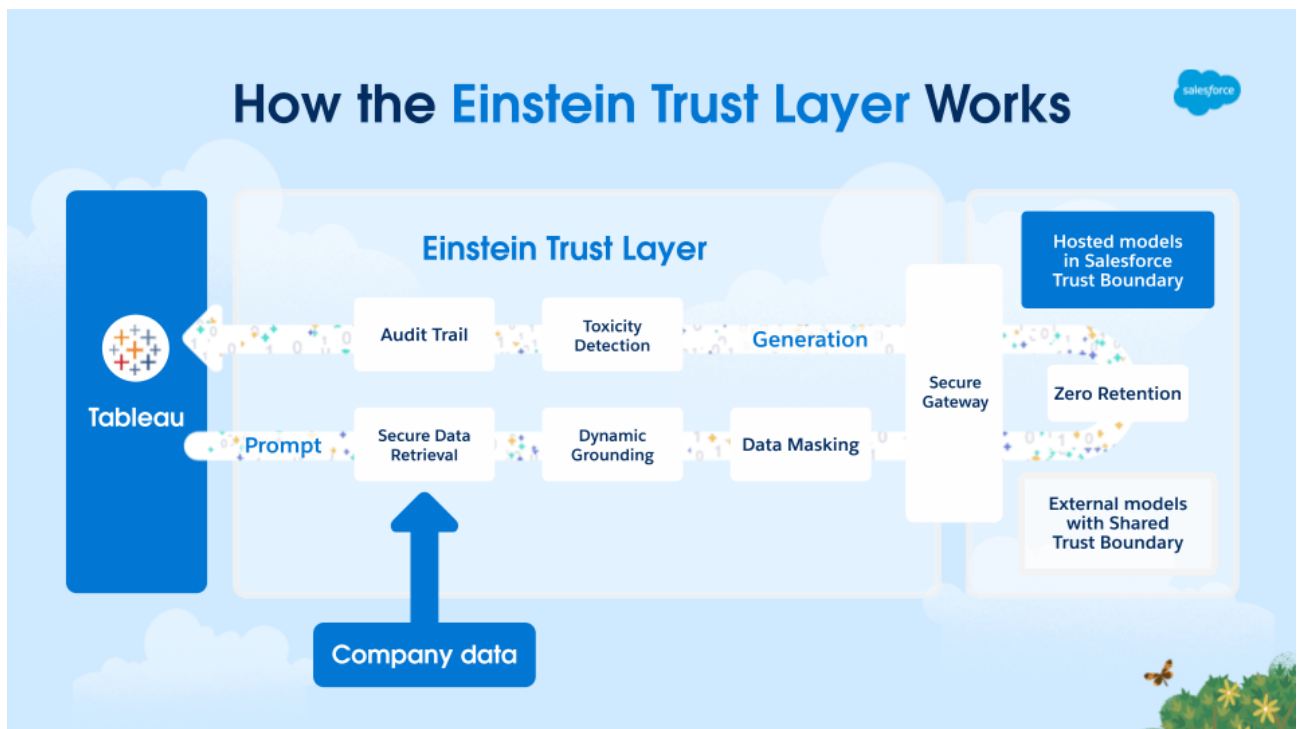


truy vấn tiếp theo và chuyển tiếp các hình ảnh hóa để hiểu đến những người có thể khai thác giá trị từ dữ liệu.

- Đảm bảo mọi người hiểu dữ liệu và có thể đưa ra quyết định sáng suốt là một trong những yếu tố quan trọng dẫn đến thành công. Tableau giúp việc cộng tác xoay quanh dữ liệu trở nên dễ dàng. Từ việc chỉnh sửa và tạo nội dung trên web đến các đề xuất nguồn dữ liệu, cung cấp cho mọi người khả năng tương tác với dữ liệu mà họ cần.
- Publishing dashboards lên Tableau Server hoặc Tableau Cloud chỉ trong vài giây, giúp mọi người trong doanh nghiệp dễ dàng thấy được những insight của vấn đề, đặt câu hỏi và đưa ra quyết định.

## 2.3 Ưu, khuyết điểm

### 2.3.1. Ưu điểm



- **Khả năng trực quan hóa dữ liệu mạnh mẽ:** Tableau cung cấp các biểu đồ và dashboard đẹp mắt, trực quan.
- **Dễ sử dụng:** Giao diện kéo và thả giúp người dùng không chuyên về kỹ thuật vẫn dễ dàng sử dụng.
- **Cộng đồng hỗ trợ lớn:** Cộng đồng người dùng đông đảo, cung cấp nhiều tài nguyên học tập và chia sẻ kinh nghiệm.
- **Tích hợp rộng rãi:** Kết nối với nhiều loại cơ sở dữ liệu, dịch vụ đám mây và các nguồn dữ liệu khác.

- **Sự hỗ trợ của AI:** Tableau có sự hỗ trợ của những model AI mới nhất như **Einstein Discovery** Được hỗ trợ bởi máy học (ML), Einstein Discovery đưa ra các dự đoán và đề xuất trong quy trình làm việc của Tableau để đưa ra quyết định thông minh hơn. Các nhà khoa học dữ liệu, nhà phân tích và người dùng doanh nghiệp đều có thể tạo ra các mô hình dự đoán mạnh mẽ mà không cần phải viết thuật toán. Được phát triển bởi Salesforce, Einstein Discovery cung cấp một môi trường trực quan, không cần mã, trao quyền cho bất kỳ ai đưa ra quyết định một cách nhanh chóng và tự tin được hướng dẫn bởi AI minh bạch, có đạo đức trong Tableau.
- **Tích hợp nhiều ngôn ngữ hỗ trợ phân tích dữ liệu** như R, Python,... giúp chúng ta có thể phân tích sâu hơn nữa vào dữ liệu, áp dụng những mô hình theo ý muốn của chúng ta.

### 2.3.2. Khuyết điểm

- **Chi phí cao:** So với các công cụ khác, chi phí của Tableau có thể khá cao.
- **Hạn chế về tùy chỉnh:** Các khả năng tùy chỉnh và tự động hóa có thể bị hạn chế so với các công cụ phân tích dữ liệu khác.
- **Yêu cầu tài nguyên:** Cần tài nguyên phần cứng đáng kể để hoạt động hiệu quả, đặc biệt khi làm việc với dữ liệu lớn.

---

## 3. Phụ lục

---

Cấu trúc bài nộp gồm 3 phần:

- Thư mục products:
  - + N01\_TLab03.twbx: File tableau chứa dashboard của nhóm
  - + products.txt: File text chứa đường link tới dashboard mà nhóm đã chia sẻ lên Tableau Server
- Thư mục docs:
  - + N01\_TLab03.pdf: File report của nhóm
- Thư mục videos:
  - + videos.txt: File text chứa đường link tới video demo của nhóm

---

## TÀI LIỆU THAM KHẢO

---

- [Khóa học đầy đủ về Tableau-Edureka](#)
- [Tableau Tutorial-By Simplilearn](#)
- [How to Create a Dashboard in Tableau - datacamp](#)
- [Tableau Blog](#)