**TRƯỜNG CAO ĐẲNG FPT POLYTECHNIC**

****

**BÁO CÁO DỰ ÁN TỐT NGHIỆP**

**Chuyên ngành Xử Lý Dữ Liệu  
<<TÊN ĐỀ TÀI>>**

Các thành viên:

* Nguyễn Văn Tý – PS00001 (Nhóm trưởng)
* Lê Thị Dậu – PS00002
* Trần Văn Ngọ – PS00003

GV HD: Ths. Nguyễn Huy Hoàng

TP.HCM 12 - 2023

MỤC LỤC

[1 Giới thiệu dự án 4](#_Toc165991907)

[1.1 Giới thiệu 4](#_Toc165991908)

[1.2 Yêu cầu của công ty 4](#_Toc165991909)

[1.3 Lập kế hoạch dự án 4](#_Toc165991910)

[2 Phân tích yêu cầu khách hàng 5](#_Toc165991911)

[2.1 Phân tích yêu cầu 5](#_Toc165991912)

[2.2 Câu chuyện dữ liệu 5](#_Toc165991913)

[2.2.1 Đặt vấn đề 5](#_Toc165991914)

[2.2.2 Xác định câu chuyện 6](#_Toc165991915)

[2.2.3 Xác định rõ đối tượng 6](#_Toc165991916)

[2.2.4 Xác định câu chuyện chi tiết 6](#_Toc165991917)

[2.2.5 Trình bày dữ liệu 6](#_Toc165991918)

[2.2.6 Những điều cần lưu ý 6](#_Toc165991919)

[2.3 Kiến trúc hệ thống 7](#_Toc165991920)

[2.3.1 Kiến trúc 7](#_Toc165991921)

[2.3.2 Giải thích 7](#_Toc165991922)

[2.4 Giải thích về bộ dữ liệu khách hàng 7](#_Toc165991923)

[2.4.1 Các khái niệm 7](#_Toc165991924)

[2.4.2 Các trường dữ liệu 7](#_Toc165991925)

[3 Làm sạch và chuyển đổi dữ liệu 8](#_Toc165991926)

[3.1 Chuẩn bị dữ liệu 8](#_Toc165991927)

[3.1.1 Giải pháp lưu trữ dữ liệu 8](#_Toc165991928)

[3.1.2 Giải pháp phân bố dữ liệu 8](#_Toc165991929)

[3.2 Làm sạch dữ liệu 8](#_Toc165991930)

[3.2.1 Các vấn đề ảnh hưởng tới dữ liệu 8](#_Toc165991931)

[3.2.2 Các tiêu chí đánh giá chất lượng dữ liệu 9](#_Toc165991932)

[3.2.3 Các bước làm sạch dữ liệu 9](#_Toc165991933)

[3.3 Chuyển đổi dữ liệu 9](#_Toc165991934)

[3.3.1 Các trường hợp cần chuyển đổi 9](#_Toc165991935)

[3.3.2 Các kỹ thuật chuyển đổi 9](#_Toc165991936)

[3.3.3 Trình bày các phép chuyển đổi trong dự án 9](#_Toc165991937)

[4 Xử lý dữ liệu 10](#_Toc165991938)

[4.1 Chuẩn hóa dữ liệu 10](#_Toc165991939)

[4.1.1 Trình bày các bước chuẩn hóa trong dự án 10](#_Toc165991940)

[4.2 Mô hình hóa dữ liệu 10](#_Toc165991941)

[4.2.1 Các loại mô hình hóa 10](#_Toc165991942)

[4.2.2 Các tiêu chí đánh giá mô hình dữ liệu 10](#_Toc165991943)

[4.2.3 Trình bày các bước mô hình hóa 10](#_Toc165991944)

[4.2.4 Trình bày các bước tạo bảng dữ liệu 11](#_Toc165991945)

[5 Trực quan hóa dữ liệu 11](#_Toc165991946)

[5.1 Các kỹ thuật trực quan hóa 11](#_Toc165991947)

[5.2 Các nguyên tắc trực quan hóa 11](#_Toc165991948)

[5.3 Trình bày tạo các report cho dự án 11](#_Toc165991949)

[5.3.1 Tạo visual thống kê chi tiết 11](#_Toc165991950)

[5.3.2 Tạo visual thống kê tổng thể 12](#_Toc165991951)

[6 Xây dựng báo cáo 12](#_Toc165991952)

[6.1 Dashboard và report 12](#_Toc165991953)

[6.2 Xây dựng báo cáo 12](#_Toc165991954)

[6.2.1 Dashboard vs Report 12](#_Toc165991955)

[6.2.2 Dashboard 13](#_Toc165991956)

[6.2.3 Report 13](#_Toc165991957)

[6.2.4 Bookmark 13](#_Toc165991958)

[7 KẾT LUẬN 14](#_Toc165991959)

[7.1 Báo cáo 14](#_Toc165991960)

[7.1.1 Các bước viết báo cáo 14](#_Toc165991961)

[7.1.2 Tổng hợp 14](#_Toc165991962)

[7.2 Khó khăn 14](#_Toc165991963)

[7.3 Thuận lợi 14](#_Toc165991964)

[7.4 Hướng phát triển 14](#_Toc165991965)

# Giới thiệu dự án

## Giới thiệu

* *Giới thiệu về hiện trạng*- Hiện trạng ….  
  - Thông tin bộ dữ liệu

## Yêu cầu của công ty

* *Nêu yêu cầu*- Về mặt dữ liệu: …  
  - Quản lý và lưu trữ: …  
  - Mục tiêu: …
* *Đánh giá tính khả thi (năng lực, khả năng của mình có thể thực hiện dự án này)*- Năng lực (skill đang có): …  
  - Năng lực (skill sẽ học thêm cho dự án): …

## Lập kế hoạch dự án

* *Lập kế hoạch kế hoạch dự án theo mẫu sau*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TT** | **Hạng mục** | **Bắt đầu** | **Kết thúc** | **Kết quả** |
| 1 | Giới thiệu dự án |  |  |  |
| 1.1 | Giới thiệu |  |  |  |
| 1.2 | Yêu cầu |  |  |  |
|  | … |  |  |  |
| 2 | Phân tích yêu cầu |  |  |  |
| 2.1 | Phân tích yêu cầu KH |  |  |  |
| 2.2 | Câu chuyện dữ liệu |  |  |  |
|  | … |  |  |  |
| 3 | Làm sạch và chuyển đổi dữ liệu |  |  |  |
| 3.1 | Chuẩn bị dữ liệu |  |  |  |
|  | … |  |  |  |
| 4 | Xử lý dữ liệu |  |  |  |
| 4.1 | Chuẩn hóa dữ liệu |  |  |  |
|  | … |  |  |  |
| 5 | Trực quan hóa dữ liệu |  |  |  |
| 5.1 | Các kĩ thuật trực quan hóa |  |  |  |
|  | … |  |  |  |
| 6 | Xây dựng báo cáo |  |  |  |
| 6.1 | Dashboard và Report |  |  |  |
|  | … |  |  |  |

# Phân tích yêu cầu khách hàng

* *Tiến hành phân tích yêu cầu khách hàng theo các bước sau đây*

## Phân tích yêu cầu

* *Dựa vào bộ dữ liệu công cty cung cấp và các yêu cầu khách hàng gửi, tiến hành phân tích:*- Dữ liệu: …  
  - Quản lý và lưu trữ: …  
  - Công nghệ: …
* *Quyết định dùng công nghệ nào*- Giải thích vì sao chọn ? Dựa vào yêu cầu nào của khách hàng ở phần trên.  
  - Dữ liệu: …  
  - Quản lý và lưu trữ: …  
  - Công nghệ: …

## Câu chuyện dữ liệu

* *Hãy mô tả các bước cụ thể để xây dựng câu chuyện, bằng cách đặt câu hỏi:*
  + *Ai đọc báo cáo này?*
  + *Giải thích vấn đề là gì? Ở mức độ nào?*
  + *Sử dụng phương pháp gì?*

### Đặt vấn đề

* **Mô tả thực trạng:**

Hiện nay ….  
……

* **Dữ liệu liên quan:**

Dữ liệu sau khi xử lý, kết quả là hàng loạt con số khô khan.  
Làm thế nào để diễn giải và truyền đạt kết quả này cho ai?  
……

* **Mục tiêu:**

Truyền đạt thông điệp dữ liệu *rõ ràng, hấp dẫn* giúp người nghe hiểu được vấn đề *nhanh chính xác*.  
Thể hiện *sinh động, dễ hiểu, thuyết phục*.  
……

### Xác định câu chuyện

* *Bằng cách đặt câu hỏi, hình thành giả thuyết, tổng hợp đào sâu dữ liệu để tìm câu trả lời:*
  + *Bạn đang cố gắng giải thích điều gì từ dữ liệu?*
  + *Mục tiêu cụ thể của bạn khi kể câu chuyện dữ liệu này là gì?*
  + *Bạn muốn đề xuất giải pháp gì từ những phân tích này?*
* *Một số cách tiếp cận dữ liệu:*
  + *Tìm kiếm mối tương quan?*
  + *Xác định xu hướng?*
  + *Rút ra so sánh?*
  + *…*

...

### Xác định rõ đối tượng

* *Trả lời các câu hỏi:*
  + *Ai cần nghe câu chuyện này?*
  + *Họ đã biết đến lĩnh vực này chưa?*
  + *….*

…

### Xác định câu chuyện chi tiết

* *Giải thích dữ liệu rõ ràng và cần trình tự logic hợp lý:*
  + *Bối cảnh?*
  + *Ai sẽ xem báo cáo này?*
  + *….*

…

### Trình bày dữ liệu

* *Dùng hình ảnh để trình bày:*
  + *Sử dụng biểu đồ gì?*
  + *….*

…

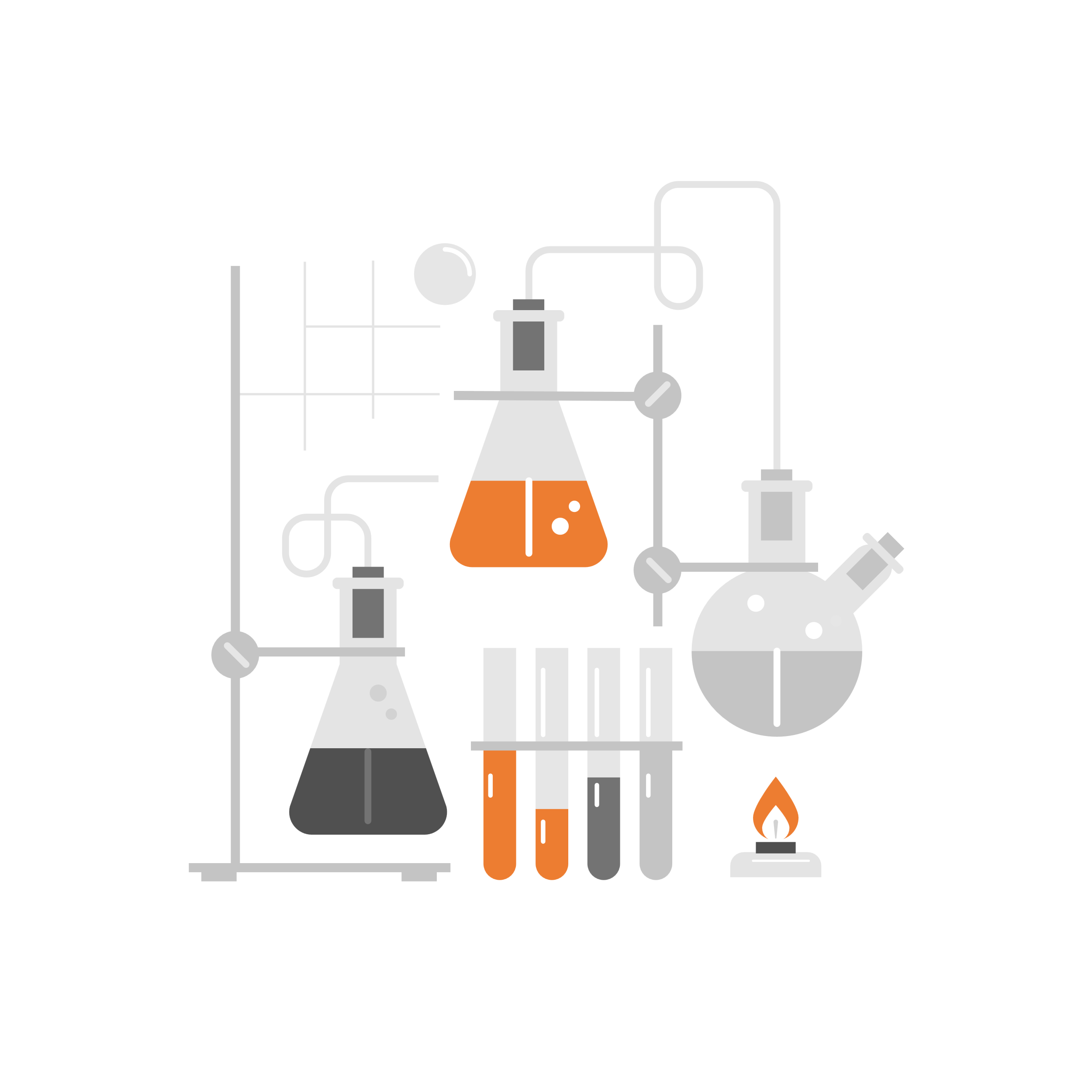
### Những điều cần lưu ý

…

## Kiến trúc hệ thống

### Kiến trúc

* *Trình bày kiến trúc hệ thống của dự án*



…

### Giải thích

* *Giải thích cách thiết kế.*

…

## Giải thích về bộ dữ liệu khách hàng

* *Cung cấp kiến thức cơ bản về bộ dữ liệu cung cấp*

### Các khái niệm

* *Giải thích các khái niệm và nghiệp vụ liên quan*

…

### Các trường dữ liệu

* *Giải thích các trường dữ liệu có trong bộ dữ liệu*

…

# Làm sạch và chuyển đổi dữ liệu

* *Bao gồm 2 bước:*
  + *Làm sạch dữ liệu*
  + *Chuyển đổi dữ liệu*

## Chuẩn bị dữ liệu

* *Quá trình chuẩn bị dữ liệu thường bao gồm việc chuẩn hóa các định dạng dữ liệu, kết hợp các bộ dữ liệu khác nhau để gia tăng độ phong phú của dữ liệu nguồn và loại bỏ các giá trị ngoại lai (outliers)*

### Giải pháp lưu trữ dữ liệu

* *So sánh giải pháp nền tảng đám mây (Cloud-based solution) hay* *các ứng dụng tại chỗ (on-premise) .*
* *Quyết định chọn giải pháp nào cho dự án? Lý do?*

…

### Giải pháp phân bố dữ liệu

* *Nhân bản là một kĩ thuật quan trọng và hữu hiệu trong việc phân bố cơ sở dữ liệu (CSDL) và thực thi các Stored procedure. Đây là chủ đề rất rộng, chỉ tập trung chọn giải pháp trong phạm vi dự án.*

#### Ý nghĩa việc phân bố dữ liệu

…

#### Trình bày cách phân bố dữ liệu

* *Trình bày cách phân bố CSDL trong phạm vi dự án này.*

…

## Làm sạch dữ liệu

* *Kỹ thuật làm sạch, tầm quan trọng, lợi ích và tiêu chí đánh giá dữ liệu.*

### Các vấn đề ảnh hưởng tới dữ liệu

* *Trình bày các vấn đề ảnh hưởng.*
* *Vấn đề nào đang tồn tại trong dự án này.*

…

### Các tiêu chí đánh giá chất lượng dữ liệu

* *Trình bày các tiêu chí.*
* *Tiêu chí nào sẽ áp dụng trong dự án này.*

…

### Các bước làm sạch dữ liệu

* *Quá trình làm sạch dữ liệu bao gồm một số bước xác định và sửa chữa các mục bị lỗi.*

#### Trình bày các bước làm sạch

…

#### Trình bày các bước làm sạch trong phạm vi dự án

* *Trình bày cụ thể các bước tác động cụ thể đến bộ dữ liệu dự án (từng cột, từng dòng).*
* *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

…

## Chuyển đổi dữ liệu

* *Các trường hợp và kỹ thuật chuyển đổi dữ liệu.*

### Các trường hợp cần chuyển đổi

* *Nêu các trường hợp cần chuyển đổi dữ liệu*

…

### Các kỹ thuật chuyển đổi

* *Trình bày các kỹ thuật chuyển đổi*
* *Kỹ thuật nào được chọn áp dụng cho dự án.*

…

### Trình bày các phép chuyển đổi trong dự án

* *Trình bày cụ thể các bước tác động cụ thể đến bộ dữ liệu dự án (từng cột, từng dòng).*
* *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

…

# Xử lý dữ liệu

## Chuẩn hóa dữ liệu

* *Chuẩn hóa dữ liệu:*
  + *Power Query*
  + *Chỉnh sửa table, column, data type*

### Trình bày các bước chuẩn hóa trong dự án

* *Trình bày cụ thể các bước chuẩn hóa dữ liệu bằng Power Query tác động cụ thể đến bộ dữ liệu dự án (từng cột, từng dòng).*
* *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

**…**

## Mô hình hóa dữ liệu

* *Mô hình hóa dữ liệu:*
  + *Data model*
  + *Data modeling*

### Các loại mô hình hóa

* *Trình bày các loại mô hình*
* *Mô hình nào đang được áp dụng cho bộ dữ liệu dự án ?*

…

### Các tiêu chí đánh giá mô hình dữ liệu

* *Trình bày các tiêu chí đánh giá mô hình dữ liệu tốt ?*
* *Mô hình đang dùng trong dự án đáp ứng các tiêu chí nào ?*

…

### Trình bày các bước mô hình hóa

* *Trình bày cụ thể các bước mô hình hóa dữ liệu bằng Power BI trên bộ dữ liệu của dự án.*
* *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

…

### Trình bày các bước tạo bảng dữ liệu

* *Trình bày cụ thể các bước tạo bảng dữ liệu bằng Power BI trên bộ dữ liệu của dự án.*
* *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

…

# Trực quan hóa dữ liệu

## Các kỹ thuật trực quan hóa

* *Trình bày các kỹ thuật trực quan hóa*
* *Kỹ thuật nào đang được áp dụng cho dự án ? Vì sao ?*

…

## Các nguyên tắc trực quan hóa

* *Trình bày các nguyên tắc trong trực quan hóa*
* *Nguyên tắc nào bạn cho rằng quan trọng nhất trong phạm vi dự án này ? Vì sao ?*

…

## Trình bày tạo các report cho dự án

* *Tạo các bộ lọc, biểu đồ theo nhu cầu dự án.*
* *Có thể phân chia theo đối tượng trong mô hình.*
* *Có thể phân chia theo chức năng hoặc page.*
* *Trong dự án cho môn học này thì phân chia theo page.*

### Tạo visual thống kê chi tiết

* *Tạo các visual dùng trong trang thống kê chi tiết.*
* *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

#### Tạo visual filter theo

…

#### Tạo visual filter theo

…

#### Tạo visual filter theo

…

#### Tạo visual tiếp theo ….

…

### Tạo visual thống kê tổng thể

* *Tạo các visual dùng trong trang thống kê tổng thể.*
* *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

#### Tạo visual filter theo ngày giao dịch

…

#### Tạo visual thống kê Tổng GTGD

…

#### Tạo visual thống kê Tổng KLGD

…

#### Tạo visual tiếp theo ….

…

# Xây dựng báo cáo

## Dashboard và report

* *Cách tối ưu hóa:*
  + *Dashboard*
  + *Report*

…

## Xây dựng báo cáo

* *Xây dựng báo cáo:*
  + *Dashboard*
  + *Report*

### Dashboard vs Report

* *Trình bày sự khác nhau giữa Dashboard và Report.*
* *Giải thích vì sao làm dashboard, report trong dự án này ?*

…

*Giải thích:* …

### Dashboard

* *Trình bày các loại dashboard có trong dự án.*
* *Giải thích vì sao làm các dashboard này ?*

…

*Giải thích:* …

### Report

* *Trình bày các loại report có trong dự án.*
* *Giải thích vì sao làm các report này ?*

#### Tạo report Detail

*Giải thích:* …

…

#### Tạo report Overview

*Giải thích:* …

…

### Bookmark

* *Trình bày các loại bookmark có trong dự án.*
* *Giải thích vì sao làm các bookmark này ?*

#### Tạo bookmark slicer

*Giải thích:* ….

#### Tạo bookmark slicer chọn

…

*Giải thích:* …

#### Tạo bookmark slicer chọn

…

*Giải thích:* …

#### Tạo visual tiếp theo ….

…

*Giải thích:* …

# KẾT LUẬN

## Báo cáo

### Các bước viết báo cáo

* *Trình bày các bước viết báo cáo phân tích dữ liệu*

…

### Tổng hợp

* *Tổng hợp các câu trả lời trong Câu Chuyện Dữ Liệu ở Phân Tích Khách Hàng, kết hợp với các bước viết báo cáo phân tích, để tạo thành bản báo cáo hoàn chỉnh.*

*Gợi ý:*

…

## Khó khăn

## Thuận lợi

## Hướng phát triển