|  |  |
| --- | --- |
|  |  |
| ĐOÀN DUY THÀNH | **BỘ CÔNG THƯƠNG**  **TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**  **---------------------------------------** |
| ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC KHOA HỌC MÁY TÍNH  **XÂY DỰNG ỨNG DỤNG LÀM SẠCH, LƯU TRỮ, PHÂN TÍCH VÀ TRỰC QUAN HÓA CÁC YẾU TỐ PHÁT TRIỂN CỦA DOANH NGHIỆP SỬ DỤNG PYTHON, MYSQL VÀ POWER BI** |
| |  |  | | --- | --- | | **CBHD:** | **ThS. Vũ Duy Giang** | | **Sinh viên:** | **Đoàn Duy Thành** | | **Mã sinh viên:** | **2020603903** |   Hà Nội – Năm 2024 |
| KHOA HỌC MÁY TÍNH |

# LỜI CẢM ƠN

Trước tiên với tình cảm sâu sắc và chân thành nhất, cho phép em được bày tỏ lòng biết ơn đến các thầy cô của trường Đại học Công Nghiệp Hà Nội, đặc biệt là các thầy cô khoa Công Nghệ Thông Tin của trường đã tạo điều kiện cho em có một kỳ đồ án tốt nghiệp.

Và em cũng xin chân thành cảm ơn thầy giáo ThS. Vũ Duy Giang người đã tận tâm hướng dẫn em qua từng buổi nói chuyện, hướng dẫn, thảo luận về các vấn đề trong đề tài. Em rất biết ơn các góp ý, sự chỉ dẫn và kiến thức chuyên môn mà thầy giáo đã chia sẻ với em trong quá trình thực hiện báo cáo.

Không chỉ hướng dẫn và hỗ trợ em hoàn thành đồ án tốt nghiệp một cách thành công, sự đóng góp của thầy giáo còn giúp em hiểu rõ hơn về lĩnh vực chuyên môn và củng cố kiến thức hoàn thành đồ án tốt nghiệp. Nhờ thầy giáo mà em đã có cơ hội nắm bắt những khía cạnh quan trọng và phát triển tư duy sáng tạo. Đồ án này không thể hoàn thành một cách xuất sắc như vậy nếu không có sự giúp đỡ của thầy. Em rất cảm ơn sự tâm huyết và nhiệt tình giúp đỡ trong xuyên suốt quá trình thực hiện báo cáo này.

Trong quá trình thực hiện báo cáo, cũng như là trong quá trình thực hiện đề tài đã chọn, em cũng đã rất cố gắng xây dựng và hoàn thiện sản phẩm. Tuy nhiên trong quá trình thực hiện sẽ khó tránh khỏi những thiết hụt và sai sót. Em rất mong nhận được ý kiến đóng góp từ thầy cô để học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt bài báo cáo tốt nghiệp. Em xin chúc thầy luôn mạnh khỏe, tràn đầy may mắn và thành công trong cuộc sống.

Em xin chân thành cảm ơn!

**Sinh viên thực hiện**

Đoàn Duy Thành

# MỤC LỤC

[LỜI CẢM ƠN ii](#_Toc176685239)

[MỤC LỤC iii](#_Toc176685240)

[DANH MỤC HÌNH ẢNH v](#_Toc176685241)

[DANH MỤC BẢNG BIỂU vii](#_Toc176685242)

[LỜI MỞ ĐẦU 1](#_Toc176685243)

[1. Lý do lựa chọn đề tài 1](#_Toc176685244)

[2. Mục tiêu của đề tài 2](#_Toc176685245)

[3. Phạm vi đề tài 3](#_Toc176685246)

[4. Dự kiến kết quả 3](#_Toc176685247)

[5. Bố cục đề tài 3](#_Toc176685248)

[CHƯƠNG 1. TỔNG QUAN VỀ ĐỀ TÀI 5](#_Toc176685249)

[1.1. Tổng quan về thị trường 5](#_Toc176685250)

[1.2. Vấn đề hiện nay 6](#_Toc176685251)

[1.3. Các mục tiêu cần giải quyết 6](#_Toc176685252)

[CHƯƠNG 2. PHÂN TÍCH THIẾT KẾ 8](#_Toc176685253)

[2.1. Khảo sát yêu cầu 8](#_Toc176685254)

[2.1.1. Mục tiêu khảo sát 8](#_Toc176685255)

[2.1.2. Kế hoạch phỏng vấn 8](#_Toc176685256)

[2.1.3. Phiếu phỏng vấn 9](#_Toc176685257)

[2.2. Phân tích yêu cầu 12](#_Toc176685258)

[2.2.1. Phân tích yêu cầu chức năng 12](#_Toc176685259)

[2.2.2. Phân tích yêu cầu phi chức năng 13](#_Toc176685260)

[2.3. Phân tích ứng dụng 15](#_Toc176685261)

[2.3.1. Hệ thống cơ sở dữ liệu 15](#_Toc176685262)

[2.3.2. Giao diện và chức năng ứng dụng 17](#_Toc176685263)

[CHƯƠNG 3. THỰC NGHIỆM 18](#_Toc176685264)

[3.1. Thiết kế cơ sở dữ liệu 18](#_Toc176685265)

[3.1.1. Biểu đồ Entity Relationship Diagram 18](#_Toc176685266)

[3.1.2. Chi tiết các bảng trong cơ sở dữ liệu 18](#_Toc176685267)

[3.2. Thiết kế giao diện ứng dụng 32](#_Toc176685268)

[3.2.1. Giao diện đăng nhập 32](#_Toc176685269)

[3.2.2. Giao diện trang chủ 32](#_Toc176685270)

[3.2.3. Giao diện thêm dữ liệu 33](#_Toc176685271)

[3.2.4. Giao diện cập nhật dữ liệu 34](#_Toc176685272)

[3.2.5. Giao diện xóa dữ liệu 35](#_Toc176685273)

[3.2.6. Giao diện làm sạch dữ liệu 36](#_Toc176685274)

[3.3. Thiết kế Dashboard 37](#_Toc176685275)

[3.3.1. Giao diện thẻ Home 37](#_Toc176685276)

[3.3.2. Giao diện thẻ HR 37](#_Toc176685277)

[3.3.3. Giao diện thẻ Marketing 38](#_Toc176685278)

[3.3.4. Giao diện thẻ Sale 39](#_Toc176685279)

[3.3.5. Giao diện thẻ Accounting 39](#_Toc176685280)

[CHƯƠNG 4: KẾT LUẬN 41](#_Toc176685281)

[TÀI LIỆU THAM KHẢO 42](#_Toc176685282)

[PHỤ LỤC 43](#_Toc176685283)

# DANH MỤC HÌNH ẢNH

[Hình 3.1: Biểu đồ Entity Relationship Diagram 18](#_Toc176685284)

[Hình 3.2: Bảng Applicant 18](#_Toc176685285)

[Hình 3.3: Bảng Application 19](#_Toc176685286)

[Hình 3.4: Bảng Asset 19](#_Toc176685287)

[Hình 3.5: Bảng Camgain 20](#_Toc176685288)

[Hình 3.6: Bảng Department 20](#_Toc176685289)

[Hình 3.7: Bảng Employee 21](#_Toc176685290)

[Hình 3.8: Bảng Employeeror 21](#_Toc176685291)

[Hình 3.9: Bảng Errorcode 21](#_Toc176685292)

[Hình 3.10: Bảng Expensereport 22](#_Toc176685293)

[Hình 3.11: Bảng Fund 22](#_Toc176685294)

[Hình 3.12: Bảng InterView 22](#_Toc176685295)

[Hình 3.13: Bảng JobPosition 23](#_Toc176685296)

[Hình 3.14: Bảng KPIAcounting 23](#_Toc176685297)

[Hình 3.15: Bảng KPIHR 24](#_Toc176685298)

[Hình 3.16: Bảng KPIMarketing 24](#_Toc176685299)

[Hình 3.17: Bảng KPISale 25](#_Toc176685300)

[Hình 3.18: Bảng Leads 25](#_Toc176685301)

[Hình 3.19: Bảng Order 26](#_Toc176685302)

[Hình 3.20: Bảng Pageview 26](#_Toc176685303)

[Hình 3.21: Bảng Payable 26](#_Toc176685304)

[Hình 3.22: Bảng Payment 27](#_Toc176685305)

[Hình 3.23: Bảng Performanceevaluationaccounting 27](#_Toc176685306)

[Hình 3.24: Bảng Performanceevaluationhr 28](#_Toc176685307)

[Hình 3.25: Bảng Performanceevaluationmkt 28](#_Toc176685308)

[Hình 3.26: Bảng Performanceevaluationsale 29](#_Toc176685309)

[Hình 3.27: Bảng personalincometax 29](#_Toc176685310)

[Hình 3.28: Bảng Product 29](#_Toc176685311)

[Hình 3.29: Bảng Reason 30](#_Toc176685312)

[Hình 3.30: Bảng Receivable 30](#_Toc176685313)

[Hình 3.31: Bảng Recruitmentchannel 30](#_Toc176685314)

[Hình 3.32: Bảng SEO 31](#_Toc176685315)

[Hình 3.33: Bảng Tax 31](#_Toc176685316)

[Hình 3.34: Bảng Taxtypedescription 31](#_Toc176685317)

[Hình 3.35: Giao diện Login 32](#_Toc176685318)

[Hình 3.36: Giao diện trang chủ 33](#_Toc176685319)

[Hình 3.37: Giao diện thêm dữ liệu 34](#_Toc176685320)

[Hình 3.38: Giao diện cập nhật dữ liệu 35](#_Toc176685321)

[Hình 3.39: Giao diện xóa dữ liệu 36](#_Toc176685322)

[Hình 3.40: Giao diện làm sạch dữ liệu 36](#_Toc176685323)

[Hình 3.41: Giao diện Dashboard thẻ Home 37](#_Toc176685324)

[Hình 3.42: Giao diện Dashboard thẻ HR 38](#_Toc176685325)

[Hình 3.43. Giao diện thẻ Marketing 38](#_Toc176685326)

[Hình 3.44: Giao diện thẻ Sale 39](#_Toc176685327)

[Hình 3.45: Giao diện thẻ Accounting 40](#_Toc176685328)

# DANH MỤC BẢNG BIỂU

[Bảng 2.1: Bảng kế hoạch phỏng vấn 8](#_Toc176685329)

[Bảng 2.2: Bảng phiếu phỏng vấn 9](#_Toc176685330)

# LỜI MỞ ĐẦU

## Lý do lựa chọn đề tài

Trong bối cảnh doanh nghiệp hiện đại ngày nay, việc quản lý và phân tích dữ liệu trở thành một yếu tố quan trọng quyết định sự thành công và phát triển bền vững. Đề tài "Xây dựng ứng dụng làm sạch, lưu trữ, phân tích và trực quan hóa các yếu tố phát triển của doanh nghiệp sử dụng Python, MySQL và Power BI" được chọn nhằm đáp ứng nhu cầu cấp thiết này.

1. **Tăng cường quản lý dữ liệu**: Doanh nghiệp ngày càng phải đối mặt với khối lượng dữ liệu khổng lồ từ nhiều nguồn khác nhau. Việc làm sạch và lưu trữ dữ liệu hiệu quả giúp đảm bảo chất lượng dữ liệu đầu vào cho các phân tích và quyết định chiến lược. Python, với các thư viện mạnh mẽ như Pandas, là công cụ lý tưởng để xử lý và làm sạch dữ liệu.
2. **Cải thiện phân tích dữ liệu**: Để đưa ra quyết định chính xác, doanh nghiệp cần phân tích dữ liệu một cách chi tiết và toàn diện. MySQL cung cấp một hệ quản trị cơ sở dữ liệu quan hệ hiệu quả để lưu trữ và quản lý dữ liệu.
3. **Trực quan hóa thông tin hiệu quả**: Dữ liệu phân tích cần được trình bày một cách trực quan để dễ dàng tiếp cận và hiểu rõ. Power BI là công cụ mạnh mẽ trong việc tạo báo cáo và bảng điều khiển tương tác, giúp các nhà quản lý và các bên liên quan nhanh chóng nắm bắt thông tin và đưa ra quyết định chiến lược.
4. **Ứng dụng thực tiễn và tiềm năng phát triển**: Đề tài này không chỉ mang lại giá trị học thuật mà còn có ứng dụng thực tiễn cao trong môi trường doanh nghiệp. Việc triển khai một hệ thống tích hợp các công cụ như Python, MySQL, và Power BI không chỉ giúp nâng cao hiệu quả quản lý dữ liệu mà còn mở ra cơ hội cải thiện quy trình làm việc và quyết định dựa trên dữ liệu.
5. **Đáp ứng xu hướng công nghệ hiện đại**: Công nghệ phân tích dữ liệu đang trở thành xu hướng chủ đạo trong ngành công nghệ thông tin. Đề tài này phản ánh sự chuyển mình của ngành công nghiệp, từ việc dựa vào trực giác sang việc sử dụng dữ liệu và phân tích để định hướng chiến lược.

.

## Mục tiêu của đề tài

* Xây dựng ứng dụng có khả năng thu thập và làm sạch dữ liệu: Phát triển một ứng dụng hỗ trợ việc thu thập và làm sạch dữ liệu một cách nhanh chóng trước khi đưa vào cơ sở dữ liệu công ty nhằm đảm bảo tính chính xác và nhất quán của dữ liệu. Sử dụng Python để xử lý dữ liệu thô, loại bỏ lỗi và cảnh báo các lỗi trước khi cập nhật dữ liệu.
* Thiết kế và triển khai cơ sở dữ liệu: Tạo ra một cơ sở dữ liệu MySQL để lưu trữ dữ liệu đã làm sạch. Cơ sở dữ liệu này sẽ bao gồm các bảng và quan hệ cần thiết để tổ chức và quản lý dữ liệu một cách hiệu quả, hỗ trợ việc truy vấn và báo cáo.
* Tạo bảng điều khiển và báo cáo trực quan: Sử dụng Power BI để thiết kế các bảng điều khiển tương tác và báo cáo trực quan. Các báo cáo và bảng điều khiển này sẽ giúp người dùng dễ dàng theo dõi và hiểu các yếu tố phát triển của doanh nghiệp thông qua các biểu đồ, đồ thị, và số liệu trực quan.
* Tích hợp và tự động hóa quy trình: Xây dựng một ứng dụng tích hợp để kết nối các thành phần làm sạch dữ liệu, lưu trữ, phân tích và trực quan hóa.

## Phạm vi đề tài

Về các khía cạnh xử lý trong đề tài, giới hạn và phạm vi sẽ tập trung xử lý các mục sau:

* + Về phạm vi công nghệ: sử dụng Python, MySQL và Power BI
  + Phạm vi chức năng: Thu thập, làm sạch, lưu trữ, phân tích và trực quan hóa dữ liệu
  + Phạm vi dữ liệu: Dữ liệu sẽ lấy từ 4 phòng ban chính của một doanh nghiệp bao gồm Sales, Marketing, Administration, Accounting.

## Dự kiến kết quả

Các kết quả dự kiến sau khi hoàn thành đề tài này bao gồm:

* Xây dựng được ứng dụng có thể thu thập, làm sạch và lưu trữ dữ liệu
* Ứng dụng có thể thao tác để quản lý cơ sở dữ liệu một cách nhanh gọn
* Thiết kế bảng điều khiển và báo cáo trực quan tự động

## Bố cục đề tài

Nội dung quyển báo cáo đồ án tốt nghiệp sẽ bao gồm các chương sau:

**Chương 1: Tổng quan về đề tài**

Trong chương 1, em sẽ nêu ra tổng quan về đề tài, các kết quả trong và ngoài nước trong lĩnh vực này.

**Chương 2: Phân tích thiết kế**

Sau khi đưa ra các khái niệm tổng quan, em sẽ thực hiện phỏng vấn khảo sát lấy mục tiêu và yêu cầu cho bài toán.

**Chương 3: Tiến hành thực nghiệm**

Ở chương này, em sẽ tiến hành thực nghiệm thiết kế cơ sở dữ liệu, ứng dụng và bảng điều khiển dựa theo các phân tích từ chương 2.

**Kết luận:**

Cuối cùng là kết luận, tổng hợp các kết quả, và hạn chế của đề tài

# TỔNG QUAN VỀ ĐỀ TÀI

## Tổng quan về thị trường

Trong thời buổi công nghệ hiện nay, chúng ta đã thực hiện được nhiều công việc với tốc độ nhanh hơn và chi phí thấp hơn nhiều so với cách thức truyền thống qua những cách thức mới dễ tiếp cận và tương tác với người dùng hơn. Chính điều này, đã thúc đẩy sự khai sinh và phát triển của nhiều lĩnh vực khiến các công ty mới ra đời ngày càng nhiều điều này giúp nâng cao và biến đổi đáng kể văn hóa, nâng cao chất lượng cuộc sống con người hiện nay.

Trong năm 2023 theo thống kê Cục Quản lý đăng ký kinh doanh cho biết, Năm nay, gần 160.000 doanh nghiệp đăng ký thành lập mới, cao gấp 1,2 lần mức bình quân giai đoạn 2017-2022 và tăng 4,6% so với dự kiến thực hiện cả năm.

Với sự gia tăng nhanh chóng của các doanh nghiệp mới và khối lượng dữ liệu mà họ tạo ra, việc quản lý và phân tích dữ liệu trở thành một thách thức quan trọng. Nhiều doanh nghiệp mới gặp khó khăn trong việc lưu trữ và phân tích dữ liệu để ra quyết định chiến lược. Sự cần thiết phải có các công cụ và giải pháp hiệu quả trong việc quản lý dữ liệu trở nên cấp bách hơn bao giờ hết.

Việc quản lý và phân tích các thông số dữ liệu là điều bắt buộc của các cấp quản lý doanh nghiệp nếu như muốn đưa ra một chiến lược dài hạn và có lợi cho công ty trong việc phát triển. Tuy nhiên việc phải xây dựng báo cáo hàng tháng một cách thủ công sẽ tạo ra một gánh nặng không cần thiết cho các nhân sự làm báo cáo, đặc biệt là việc xây dựng một báo cáo đủ thông tin sẽ rất tốn thời gian và công sức một cách vô nghĩa trong thời đại hiện nay.

Chính vì vậy, đối với một công ty, việc lưu trữ và trực quan hóa dữ liệu của công ty một cách tự động, đầy đủ và chính xác là một việc rất cần thiết. Và đó cũng chính là mục tiêu của đề tài “Xây dựng ứng dụng làm sạch, lưu trữ, phân tích và trực quan hóa các yếu tố phát triển của doanh nghiệp sử dụng Python, MySQL và Power BI”.

## Vấn đề hiện nay

Trong nước, nhiều doanh nghiệp tại Việt Nam đã bước đầu ứng dụng các công cụ phân tích và trực quan hóa dữ liệu như Power BI, Tableau để hỗ trợ quá trình ra quyết định kinh doanh. Điển hình, các tập đoàn lớn như VinGroup, Vietcombank đã triển khai các hệ thống phân tích dữ liệu để tối ưu hóa hoạt động và gia tăng hiệu suất kinh doanh. Tuy nhiên, ở các doanh nghiệp vừa và nhỏ, việc tích hợp và làm sạch dữ liệu vẫn còn gặp nhiều khó khăn do hạn chế về hạ tầng công nghệ và nguồn lực chuyên môn.

Ngoài ra đối với các doanh nghiệp vừa và nhỏ, việc lưu trữ chủ yếu bằng excel, dẫn đến việc hạn chế khi kết nối truy vấn thông tin giữa các bảng.

Các ứng dụng cho việc trực quan một cách tự động hiện nay đang chưa có, một số doanh nghiệp Việt Nam đang phát triển về việc thiết kế các báo cáo cho các doanh nghiệp vừa và nhỏ như BASE, INDA…

Việc phát triển một ứng dụng có khả năng chỉ cần nhập dữ liệu để lưu trữ và trực quan hóa dữ liệu hiện nay là chưa có trên thị trường hiện nay.

### Các mục tiêu cần giải quyết

Các vấn đề mà đề tài đồ án cần tập trung nghiên cứu giải quyết bao gồm:

1. **Làm sạch dữ liệu**: Quá trình làm sạch dữ liệu là bước cần thiết để loại bỏ các giá trị thiếu, sai lệch hoặc không phù hợp trong dữ liệu.
2. **Lưu trữ dữ liệu**: Thiết kế và tối ưu hóa hệ thống cơ sở dữ liệu (MySQL) để lưu trữ hiệu quả khối lượng lớn dữ liệu doanh nghiệp. Cần giải quyết các vấn đề về khả năng mở rộng, an toàn và truy xuất dữ liệu.
3. **Phân tích dữ liệu**: Đề tài cần tập trung vào các phương pháp phân tích dữ liệu phù hợp với yêu cầu kinh doanh, xây dựng các DAX tính toán các thông số phù hợp.
4. **Trực quan hóa dữ liệu**: Sử dụng Power BI để xây dựng các báo cáo và biểu đồ trực quan, giúp doanh nghiệp dễ dàng nắm bắt thông tin quan trọng. Cần giải quyết các vấn đề liên quan đến tính hiệu quả, trực quan và khả năng tương tác của các báo cáo.
5. **Tích hợp hệ thống**: Đảm bảo khả năng kết nối và tích hợp giữa các thành phần trong hệ thống, bao gồm Python, MySQL và Power BI, để đảm bảo luồng dữ liệu thông suốt và liền mạch.

# PHÂN TÍCH THIẾT KẾ

## Khảo sát yêu cầu

### Mục tiêu khảo sát

Mục tiêu của việc thu thập kháo sát nhằm phân tích, tìm kiếm các chức năng mà doanh nghiệp trông đợi để giải quyết các vấn đề và bài toán của doanh nghiệp đang gặp phải liên quan đến các bất cập cần được giải quyết trong việc lưu trữ và trực quan hóa dữ liệu của công ty.

### Kế hoạch phỏng vấn

Bảng 2.1: Bảng kế hoạch phỏng vấn

|  |  |
| --- | --- |
| Người được hỏi: CEO công ty Kin Logistic | Người phỏng vấn: Đoàn Duy Thành |
| Địa chỉ: Tầng 4 số nhà 19 ngõ 68 Lưu Hữu Phước, Cầu Diễn, Nam Từ Liêm, Hà Nội | Thời gian hẹn: 19/7/2024  Thời điểm bắt đầu: 8h  Thời gian kết thúc: 9h |
| Đối tượng: Quản lý công ty  Mục tiêu phỏng vấn:   * Tìm hiểu về các yêu cầu của công ty về việc lưu trữ và trực quan hóa dữ liệu * Phân tích bài toán nghiệp vụ của từng phòng ban * Xác định các chức năng và kết quả cần đạt của ứng dụng | Yêu cầu:  Người được hỏi phải là quản lý liên quan đến việc phân tích hiệu suất công ty để đưa chiến lược cho công ty |
| Chương trình:   * Giới thiệu * Tổng quan về dự án * Tổng quan về phỏng vấn * Đâu là những vấn đề hiện có của quý công ty? * Anh mong chờ ứng dụng sẽ đạt được những gì cho các vấn đề trên? * Các nghiệp vụ chính của các phòng ban chính trong công ty là gì. * Các thông số chính trên báo cáo trực quan của từng phòng ban là gì * Trao đổi làm rõ các thắc mắc * Kết thúc (Thỏa thuận ) | Ước lượng thời gian  1 phút  2 phút  2 phút  5 phút  5 phút  15 phút  15 phút  10 phút |
|  | Dự kiến: 55 phút |

### Phiếu phỏng vấn

Bảng 2.2: Bảng phiếu phỏng vấn

|  |  |
| --- | --- |
| **PHIẾU PHỎNG VẤN** | |
| Người được hỏi: Mr. Hoàng Nhật Long | Người hỏi: Đoàn Duy Thành  Ngày: 19/07/2024 |
| **Câu hỏi** | **Ghi chú** |
| Câu 1: Đâu là những vấn đề hiện có của quý công ty? | Trả lời:   * Chưa có cơ sở dữ liệu * Thời gian làm các báo cáo định kỳ tương đối tốn thời gian do năng lực nhân sự |
| Câu 2: Anh mong chờ ứng dụng sẽ đạt được những gì cho các vấn đề trên | Trả lời:   * Ứng dụng có thể kiểm tra một số lỗi dữ liệu như thiếu, trùng lặp… * Có khả năng lưu trữ được dữ liệu của công ty một cách hiệu quả * Có tính năng thêm, sửa, xóa và báo lỗi nếu có tồn tại trong dữ liệu * Có thể báo cáo tự động với đủ insight, tham số cần thiết |
| Câu 3: Các nghiệp vụ chính của các phòng ban là gì | Trả lời:   * Phòng HR sẽ chịu trách nhiệm liên quan về mặt con người * Phòng Marketing chịu trách nhiệm về mặt truyền thông, quảng bá * Phòng Sales chịu trách nhiệm cho việc kinh doanh của công ty * Phòng kế toán sẽ chịu trách nhiệm liên quan đến các vấn đề tài chính của công ty |
| Câu 4: Các thông số chính trên báo cáo trực quan của từng phòng ban là gì | Trả lời:  Đối với HR   * Số lượng nhân sự * Tỷ lệ nghỉ việc * Số lượng CV apply * Thông số về CS và WPS (Điểm hành vi và năng lực) * Phân tích các lỗi của nhân sự công ty   Đối với Marketing:   * Chỉ số leads, CLR, CTR, Rank key. * Thông số về CS và WPS * Phân loại khách hàng   Đối với Sale:   * Doanh thu, số lượng khách hàng mới, * Số lượng đơn đặt hàng * Chỉ số ARPC * Phân tích sản phẩm yêu thích * Thông số CS và WPS   Đối với Accounting:   * Tổng doanh thu, lợi nhuận ròng, chỉ số quỹ * Tổng lượng chi tiêu * Thông số CS và WPS * Tổng vay, tổng nợ, hệ số DER |

## Phân tích yêu cầu

### Phân tích yêu cầu chức năng

Ứng dụng được phát triển nhằm mục tiêu hỗ trợ các doanh nghiệp trong việc quản lý và phân tích các yếu tố phát triển, từ đó đưa ra các quyết định chiến lược. Để đạt được mục tiêu này, ứng dụng cần đáp ứng các yêu cầu chức năng sau:

1. **Chức năng làm sạch dữ liệu (Data Cleaning):**
   * **Mô tả:** Ứng dụng cần có khả năng tự động nhận diện và loại bỏ các dữ liệu thiếu sót, dữ liệu trùng lặp, và các lỗi dữ liệu phổ biến khác.
   * **Yêu cầu:**
     + Tích hợp các thuật toán làm sạch dữ liệu hiệu quả.
     + Báo cáo phản hồi trong trường hợp dữ liệu bị lỗi
     + Cung cấp giao diện để người dùng kiểm tra làm sạch dữ liệu trước khi xác nhận nhập vào
2. **Chức năng lưu trữ dữ liệu (Data Storage):**
   * **Mô tả:** Dữ liệu sau khi được làm sạch cần được lưu trữ an toàn và có cấu trúc trong hệ quản trị cơ sở dữ liệu MySQL.
   * **Yêu cầu:**
     + Cơ sở dữ liệu cần thiết kế để đảm bảo tính toàn vẹn và nhất quán của dữ liệu.
     + Hỗ trợ truy vấn nhanh chóng và hiệu quả để phục vụ quá trình phân tích dữ liệu.
3. **Chức năng trực quan hóa dữ liệu (Data Visualization):**
   * **Mô tả:** Sau khi dữ liệu được phân tích, ứng dụng cần có khả năng trực quan hóa kết quả phân tích thông qua các báo cáo động và biểu đồ.
   * **Yêu cầu:**
     + Tích hợp Power BI để tạo các báo cáo và dashboard trực quan, dễ hiểu.
     + Cung cấp khả năng tương tác với các báo cáo, cho phép người dùng lọc và tùy chỉnh các biểu đồ theo nhu cầu.Phân tích yêu cầu phi chức năng

### Phân tích yêu cầu phi chức năng

Các yêu cầu phi chức năng đề cập đến các tiêu chí mà hệ thống phải tuân thủ để đảm bảo hoạt động hiệu quả, bảo mật, và tương tác tốt với người dùng. Dưới đây là các yêu cầu phi chức năng chính:

1. **Hiệu năng (Performance):**
   * **Mô tả:** Ứng dụng cần đảm bảo khả năng xử lý và phân tích dữ liệu một cách nhanh chóng và hiệu quả.
   * **Yêu cầu:**
     + Thời gian phản hồi cho các truy vấn tính theo giây đối với các tập dữ liệu thông thường.
2. **Khả năng mở rộng (Scalability):**
   * **Mô tả:** Hệ thống cần được thiết kế để có thể dễ dàng mở rộng nếu cần phát triển thêm.
   * **Yêu cầu:**
     + Hệ thống phải có kiến trúc cho phép dễ dàng mở rộng hạ tầng cơ sở và chức năng.
3. **Tính bảo mật (Security):**
   * **Mô tả:** Bảo mật dữ liệu là yếu tố quan trọng đối với ứng dụng, đặc biệt là đối với thông tin nhạy cảm của doanh nghiệp.
   * **Yêu cầu:**
     + Xác thực và phân quyền người dùng chi tiết để hạn chế truy cập trái phép.
4. **Khả năng tương thích (Compatibility):**
   * **Mô tả:** Ứng dụng cần tương thích với nhiều nền tảng và công nghệ khác nhau, cũng như các hệ thống hiện có của doanh nghiệp.
   * **Yêu cầu:**
     + Hỗ trợ nhiều hệ điều hành (Windows, macOS, Linux).
     + Tương thích với các phiên bản khác nhau của MySQL và các hệ thống cơ sở dữ liệu tương tự.
     + Tích hợp dễ dàng với các công cụ BI khác ngoài Power BI nếu cần thiết.
5. **Khả năng sử dụng (Usability):**
   * **Mô tả:** Ứng dụng cần có giao diện thân thiện với người dùng, dễ sử dụng và có trải nghiệm người dùng tốt.
   * **Yêu cầu:**
     + Giao diện trực quan, dễ hiểu, và dễ sử dụng cho cả người dùng không có nền tảng kỹ thuật.
     + Cung cấp tài liệu hướng dẫn sử dụng chi tiết và hỗ trợ người dùng.
6. **Khả năng bảo trì (Maintainability):**
   * **Mô tả:** Ứng dụng cần được thiết kế để dễ dàng bảo trì và cập nhật trong tương lai.
   * **Yêu cầu:**
     + Mã nguồn phải được viết rõ ràng, có chú thích và tuân theo các tiêu chuẩn lập trình.
     + Cung cấp tài liệu kỹ thuật chi tiết để hỗ trợ quá trình bảo trì và phát triển thêm.

## Phân tích ứng dụng

### Hệ thống cơ sở dữ liệu

Dựa theo phạm vi đề tài, hệ thống cơ sở dữ liệu được phục vụ để lưu trữ cho 4 phòng ban chính bao gồm Sales, Marketing, Administration và Accounting với các yêu cầu sau.

#### Sales

Phòng Sales sẽ cần quản lý và trực quan hóa các thông tin quan trọng sau để theo dõi hiệu suất và đưa ra các quyết định kinh doanh:

1. Doanh số bán hàng (Sales Revenue)
2. Số lượng đơn hàng (Number of Orders)
3. Tỷ lệ chuyển đổi (Conversion Rate)
4. Hiệu suất của nhân viên bán hàng (Salesperson Performance)
5. Sản phẩm bán chạy (Top-Selling Products)

#### Marketing

Phòng Marketing sẽ cần quản lý và trực quan hóa các thông tin quan trọng sau để theo dõi hiệu suất và đưa ra các quyết định quảng cáo:

1. Tổng số khách hàng mới
2. Tỷ lệ CLR (Conversion Lead Rate)
3. Tỷ lệ CTR (Click To Reach)
4. Xếp hạng Key
5. Phân loại khách hàng

#### HR

Phòng HR sẽ cần quản lý và trực quan hóa các thông tin quan trọng sau để theo dõi hiệu suất và đưa ra các quyết định nhân sự:

1. Tổng lượng nhân sự làm việc
2. Tỷ lệ nghỉ việc
3. Số lượng đơn ứng tuyển
4. Tỷ lệ phỏng vấn thành công
5. Điểm AWPS và ACS
6. Lỗi kỷ luật

#### Accounting

Phòng Accounting sẽ cần quản lý và trực quan hóa các thông tin quan trọng sau để theo dõi hiệu suất và đưa ra các quyết định tài chính:

1. Tổng doanh thu
2. Lợi nhuận ròng
3. Quỹ công ty
4. Thuế
5. Tỷ lệ công nợ
6. Hệ số DER và ROI

### Giao diện và chức năng ứng dụng

Giao diện của ứng dụng bao gồm:

* Giao diện đăng nhập
* Giao diện trang chủ
* Giao diện thêm dữ liệu
* Giao diện cập nhật dữ liệu
* Giao diện xóa dữ liệu
* Giao diện làm sạch dữ liệu

Chức năng ứng dụng bao gồm:

* Đăng nhập
* Thêm mới dữ liệu vào cơ sở dữ liệu
* Cập nhật dữ liệu vào cơ sở dữ liệu
* Xóa dữ liệu khỏi cơ sở dữ liệu
* Làm sạch dữ liệu

# CHƯƠNG 3. THỰC NGHIỆM

## 3.1. Thiết kế cơ sở dữ liệu

### 3.1.1. Biểu đồ Entity Relationship Diagram

A computer screen shot of a computer

Description automatically generated

Hình 3.1: Biểu đồ Entity Relationship Diagram

### 3.1.2. Chi tiết các bảng trong cơ sở dữ liệu

1. Bảng Applicant

A screenshot of a computer

Description automatically generated

Hình 3.2: Bảng Applicant

1. Bảng Application

A screenshot of a computer

Description automatically generated

Hình 3.3: Bảng Application

1. Bảng Asset

A screenshot of a computer

Description automatically generated

Hình 3.4: Bảng Asset

1. Bảng Camgain

A screenshot of a computer

Description automatically generated

Hình 3.5: Bảng Camgain

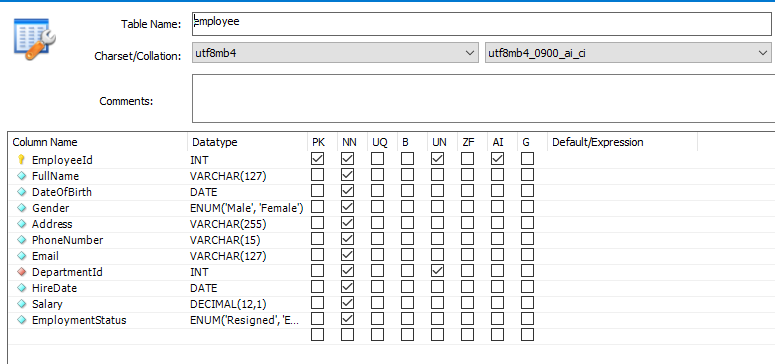
1. Bảng Department

A screenshot of a computer

Description automatically generated

Hình 3.6: Bảng Department

1. Bảng Employee



Hình 3.7: Bảng Employee

1. Bảng Employeeror

A screenshot of a computer

Description automatically generated

Hình 3.8: Bảng Employeeror

1. Bảng Errorcode

A screenshot of a computer

Description automatically generated

Hình 3.9: Bảng Errorcode

1. Bảng Expensereport

A screenshot of a survey

Description automatically generated

Hình 3.10: Bảng Expensereport

1. Bảng Fund

A screenshot of a computer

Description automatically generated

Hình 3.11: Bảng Fund

1. Bảng Interview

A screenshot of a computer

Description automatically generated

Hình 3.12: Bảng InterView

1. Bảng Job Position

A screenshot of a computer

Description automatically generated

Hình 3.13: Bảng JobPosition

1. Bảng KPIAcounting

A screenshot of a computer

Description automatically generated

Hình 3.14: Bảng KPIAcounting

1. Bảng KPIHR

A screenshot of a computer

Description automatically generated

Hình 3.15: Bảng KPIHR

1. Bảng KPIMarketing

A screenshot of a computer

Description automatically generated

Hình 3.16: Bảng KPIMarketing

1. Bảng KPISale

A screenshot of a computer

Description automatically generated

Hình 3.17: Bảng KPISale

1. Bảng Leads

A screenshot of a computer

Description automatically generated

Hình 3.18: Bảng Leads

1. Bảng Order

A screenshot of a computer

Description automatically generated

Hình 3.19: Bảng Order

1. Bảng Pageview

A screenshot of a computer

Description automatically generated

Hình 3.20: Bảng Pageview

1. Bảng Payable

A screenshot of a computer

Description automatically generated

Hình 3.21: Bảng Payable

1. Bảng payment

A screenshot of a computer

Description automatically generated

Hình 3.22: Bảng Payment

1. Bảng Performanceevaluationaccounting

A screenshot of a computer

Description automatically generated

Hình 3.23: Bảng Performanceevaluationaccounting

1. Bảng performanceevaluationhr

A screenshot of a computer

Description automatically generated

Hình 3.24: Bảng Performanceevaluationhr

1. Bảng performanceevaluationmkt

A screenshot of a computer

Description automatically generated

Hình 3.25: Bảng Performanceevaluationmkt

1. Bảng performanceevaluationsale

A screenshot of a computer

Description automatically generated

Hình 3.26: Bảng Performanceevaluationsale

1. Bảng personalincometax

A screenshot of a computer

Description automatically generated

Hình 3.27: Bảng personalincometax

1. Bảng Product

A screenshot of a computer

Description automatically generated

Hình 3.28: Bảng Product

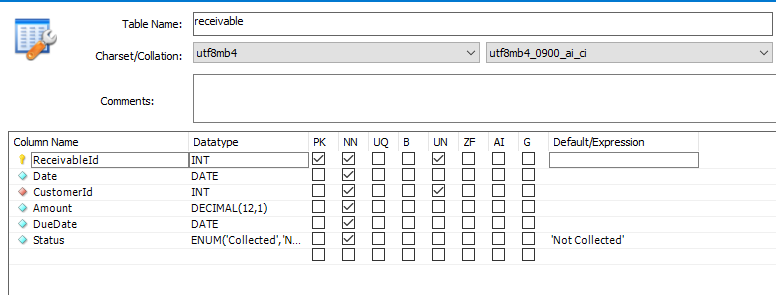
1. Bảng Reason

A screenshot of a computer

Description automatically generated

Hình 3.29: Bảng Reason

1. Bảng receivable



Hình 3.30: Bảng Receivable

1. Bảng recruitmentchannel

A screenshot of a computer

Description automatically generated

Hình 3.31: Bảng Recruitmentchannel

1. Bảng SEO

A screenshot of a computer

Description automatically generated

Hình 3.32: Bảng SEO

1. Bảng Tax

A screenshot of a computer

Description automatically generated

Hình 3.33: Bảng Tax

1. Bảng Taxtypedescription

A screenshot of a computer

Description automatically generated

Hình 3.34: Bảng Taxtypedescription

## 3.2. Thiết kế giao diện ứng dụng

### 3.2.1. Giao diện đăng nhập

Người dùng sau khi nhập tài khoản và mật khẩu hợp lệ ấn nút đăng nhập để đăng nhập ứng dụng.

A screenshot of a login screen

Description automatically generated

Hình 3.35: Giao diện Login

### 3.2.2. Giao diện trang chủ

Người dùng sau khi đăng nhập thành công sẽ được đưa đến trang chủ, tại đây người dùng sẽ lựa chọn tính năng mong muốn như thêm, sửa hoặc xóa dữ liệu tới database.

A screenshot of a computer

Description automatically generated

Hình 3.36: Giao diện trang chủ

### 3.2.3. Giao diện thêm dữ liệu

Người dùng chọn vào nút Import Data để chọn file dữ liệu muốn nhập, sau đó lựa chọn phòng ban và bảng theo mục tiêu cần lưu trữ. Sau khi hoàn tất nhấn nút Process để thực hiện thêm dữ liệu vào cơ sở dữ liệu.

Nếu đã hoàn tất có thể chọn các phòng ban khác hoặc nút mũi tên Back để quay về trang chủ.

A screenshot of a computer

Description automatically generated

Hình 3.37: Giao diện thêm dữ liệu

### 3.2.4. Giao diện cập nhật dữ liệu

Người dùng chọn vào nút Import Data để chọn file dữ liệu muốn cập nhật, sau đó lựa chọn phòng ban và bảng theo mục tiêu cần cập nhật. Sau khi hoàn tất nhấn nút Process để thực hiện cập nhật dữ liệu vào cơ sở dữ liệu.

Ngoài ra người dùng có thể nhập trực tiếp mã ID của bản ghi muốn cập nhật và sửa trực tiếp trên giao diện ứng dụng, sau khi hoàn tất nhấn nút Process để thực hiện cập nhật dữ liệu vào cơ sở dữ liệu.

Nếu đã hoàn tất có thể chọn các phòng ban khác hoặc nút mũi tên Back để quay về trang chủ.

A screenshot of a computer

Description automatically generated

Hình 3.38: Giao diện cập nhật dữ liệu

### 3.2.5. Giao diện xóa dữ liệu

Người dùng chọn vào nút Import Data để chọn file dữ liệu muốn xóa, sau đó lựa chọn phòng ban và bảng theo mục tiêu cần xóa. Sau khi hoàn tất nhấn nút Process để thực hiện xóa dữ liệu khỏi cơ sở dữ liệu.

Ngoài ra người dùng có thể nhập trực tiếp mã ID của bản ghi muốn xóa, sau khi hoàn tất nhấn nút Process để thực hiện cập nhật dữ liệu vào cơ sở dữ liệu.

Nếu đã hoàn tất có thể chọn các phòng ban khác hoặc nút mũi tên Back để quay về trang chủ.

A screenshot of a computer

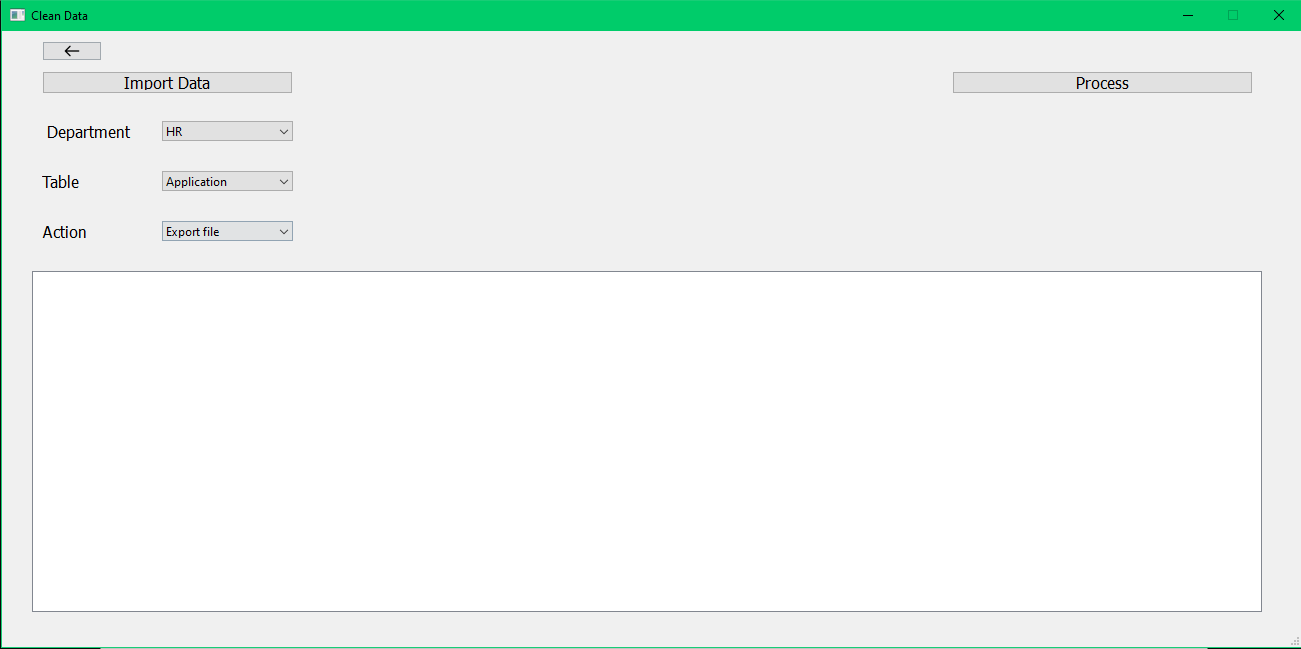
Description automatically generated

Hình 3.39: Giao diện xóa dữ liệu

### 3.2.6. Giao diện làm sạch dữ liệu

Người dùng chọn vào nút Import Data để chọn file dữ liệu muốn làm sạch, sau đó lựa chọn phòng ban và bảng theo mục tiêu cần sửa. Ngoài ra sẽ lựa chọn các hành động cho việc làm sạch.

Nếu đã hoàn tất có thể chọn các bộ dữ liệu khác hoặc nút mũi tên Back để quay về trang chủ.



Hình 3.40: Giao diện làm sạch dữ liệu

## 3.3. Thiết kế Dashboard

Sau khi đã có dữ liệu trong cơ sở dữ liệu, ứng dụng sẽ tự động tính toán và phân tích theo các DAX đã được tính toán trên nền tảng Power BI, sau đó sẽ đưa ra các thông số insight thích hợp để phục vụ cho việc phân tích.

### 3.3.1. Giao diện thẻ Home

Giao diện thẻ Home trực quan đưa ra các thông số liên quan đến doanh thu, chi phí, số tiền nợ hoặc vay cùng các chỉ số quan trọng như ROI và DER.

A screenshot of a computer

Description automatically generated

Hình 3.41: Giao diện Dashboard thẻ Home

### 3.3.2. Giao diện thẻ HR

Giao diện của thẻ HR trực quan đưa ra các thông số liên quan đến tình hình nhân sự tại công ty như số lượng nhân sự, tỷ lệ nghỉ việc, số lượng đơn ứng tuyển, hệ số AWPS và ACS…

A screenshot of a computer

Description automatically generated

Hình 3.42: Giao diện Dashboard thẻ HR

### 3.3.3. Giao diện thẻ Marketing

Giao diện của thẻ Marketing trực quan đưa ra các thông số liên quan về quảng bá thương hiệu và tiếp cận khách hàng. Một số chỉ số chính bao gồm Leads, CLR, CTR, Rank Key cùng các chỉ số nhân sự như AWPS và ACS

A screenshot of a computer

Description automatically generated

Hình 3.43. Giao diện thẻ Marketing

### 3.3.4. Giao diện thẻ Sale

Giao diện của thẻ Sale trực quan đưa ra các thông số liên quan về việc kinh doanh của doanh nghiệp bao gồm các chỉ số chính như Revenue, Customer, Order, ARPC cùng các chỉ số như AWPS và ACS.

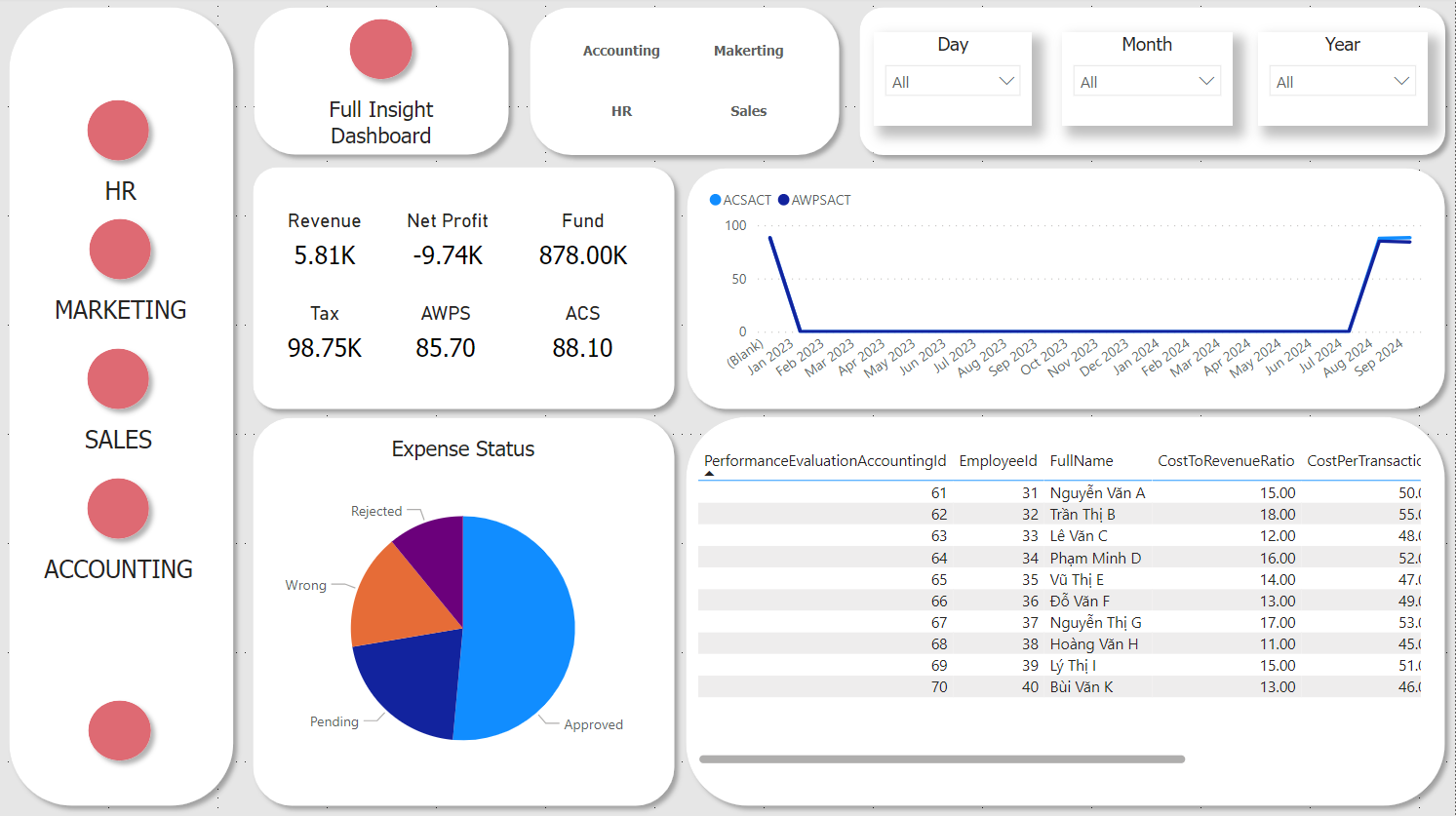
A screenshot of a computer

Description automatically generated

Hình 3.44: Giao diện thẻ Sale

### 3.3.5. Giao diện thẻ Accounting

Giao diện của thẻ Accounting trực quan đưa ra các thông số liên quan về kinh tế của doanh nghiệp bao gồm các chỉ số chính như Revenue, Net Profit, Fund, Tax cùng các chỉ số như AWPS và ACS.



Hình 3.45: Giao diện thẻ Accounting

# CHƯƠNG 4: KẾT LUẬN

Trong suốt quá trình thực hiện đồ án với đề tài“Xây dựng ứng dụng làm sạch, lưu trữ, phân tích và trực quan hóa các yếu tố phát triển của doanh nghiệp sử dụng Python, MySQL và Power BI”, em đã đạt được nhiều kinh nghiệm, trải nghiệm quý báu và tích lũy cho mình được nhiều bài học quý giá.

Đồ án đã được thực hiện thành công, đạt được mục đích và yêu cầu đặt ra. Cụ thể, đồ án đã xây dựng thành công một ứng dụng đáp ứng được các yêu cầu về chức năng, hiệu suất đồng thời cũng đạt được các kết quả sau:

* Tìm hiểu về cách xây dựng một hệ thống cơ sở dữ liệu.
* Xây dựng được một hệ thống ứng dụng có thể giúp người dùng dễ dàng quản lý cơ sở dữ liệu.
* Thiết kế thành công dashboard tự động cho việc trực quan và phân tích các yếu tố doanh nghiệp

Tuy rằng em đã xây dựng được một hệ thống ưng dụng cơ bản và hoạt động ổn định, tuy nhiên nhiên vẫn khó tránh khỏi còn một số hạn chế như sau:

* Giao diện người dùng còn đơn giản và có thể chưa đẹp mắt
* Một vài những tính năng vẫn còn chưa tối ưu hóa một cách hoàn chỉnh.

Dù đã cố gắng hết sức, nhưng với giới hạn về năng lực và kiến thức, em nhận thức rõ rằng còn nhiều hành trình phía trước để nghiên cứu và phát triển. Những hạn chế đã được liệt kê ở trên cũng chính là phương hướng giúp em phát triển trong tương lai phía trước trên con đường của mình.

# TÀI LIỆU THAM KHẢO

1. . Phạm Quý Dương, Tìm hiểu về MySQL, Viblo, năm 2021. <https://viblo.asia/p/tim-hieu-ve-my-sql-gDVK2Ov2ZLj>. [Truy cập vào 7/7/2024].
2. . Lê Hoàng Hạnh, Power BI là gì, BaseVN, năm 2024. <https://base.vn/blog/power-bi-la-gi/>. [Truy cập ngày 7/7/2024].
3. . Việt Linh, Một năm có tới gần 160.000 doanh nghiệp mới, Báo Tiên Phong, năm 2023. <https://tienphong.vn/mot-nam-co-toi-gan-160000-doanh-nghiep-moi-post1599436.tpo>. [Truy cập ngày 10/7/2024].
4. . Martin Fitzpatrick, PyQt6 Tutorial, Pythonguis, năm 2024. <https://www.pythonguis.com/pyqt6-tutorial/>. [Truy cập ngày 12/7/2024]

# PHỤ LỤC

Source code: <https://github.com/DoanDuyThanh71/ToolControlDatabase>