**TRƯỜNG ĐẠI HỌC KINH TẾ**

**KHOA THỐNG KÊ - TIN HỌC**



**BÁO CÁO THỰC TẬP NGHỀ NGHIỆP**

**TÊN ĐỀ TÀI:**

**THỰC HIỆN TRỰC QUAN HÓA DỮ LIỆU**

**CHO KHÁCH HÀNG**

**Đơn vị thực tập: Công ty cổ phần công nghệ dữ liệu DSA**

Giảng viên hướng dẫn: Cao Thị Nhâm

Sinh viên thực hiện: Nguyễn Hữu Hiếu Thiên

Lớp: 44K21.2

**Đà Nẵng, 09/07/2021**

# **LỜI CẢM ƠN**

Để hoàn thành chuyên đề báo cáo thực tập này trước hết em xin gửi đén quý thầy cô giáo trong khoa Thống Kê – Tin học trường Đại học Kinh tế Đại học Đà Nẵng lời cám ơn chân thành.

Đặc biệt, em xin gửi đến cô Cao Thị Nhâm, người đã tận tình hướng dẫn, giúp đỡ em hoàn thành chuyên đề báo cáo thực tập này lời cảm ơn sâu sắc nhất.

Em xin chân thành cảm ơn Ban Lãnh Đạo, đặc biệt là anh Trần Thiên Quốc Tổng đã tạo điều kiện thuận lợi cho bọn em được tìm hiểu thực tiễn trong suốt quá trình thực tập tại công ty.

Cuối cùng em xin cảm ơn các anh chị trong công ty cổ phần công nghệ dữ liệu DSA đã giúp đỡ, cung cấp những số liệu thực tế để em hoàn thành tốt chuyên đề thực tập nghề nghiệp này.

# **LỜI CAM ĐOAN**

Chúng em xin cam đoan dự án THỰC HIỆN TRỰC QUAN HÓA DỮ LIỆU CHO KHÁCH HÀNG là kết quả nghiên cứu độc lập dưới sự hướng dẫn của giáo viên hướng dẫn Cao Thị Nhâm và mentor Trần Thiên Quốc Tổng. Ngoài ra không có bất cứ sự sao chép nào của người khác.

Dự án, nội dung báo cáo là sản phẩm mà chúng em đã nỗ lực nghiên cứu trong quá trình học tập cũng như tìm hiểu trên ở nhiều phương tiện. Các dữ liệu, kết quả trình bày trong báo cáo là hoàn toàn trung thực. Chúng em xin cam kết đề cương chưa từng được công bố trước đây. Chúng em xin chịu hoàn toàn trách nhiệm, kỷ luật của Bộ môn và Nhà trường đề ra nếu có vi phạm.

**MỤC LỤC**

[**LỜI CẢM ƠN** 2](#_Toc76762874)

[**LỜI CAM ĐOAN** 3](#_Toc76762875)

[**DANH MỤC HÌNH ẢNH** 5](#_Toc76762876)

[**DANH MỤC CÁC TỪ VIẾT TẮT** 6](#_Toc76762877)

[LỜI MỞ ĐẦU 7](#_Toc76762878)

[**1.** **Mục tiêu nghiên cứu đề tài** 7](#_Toc76762879)

[**2.** **Nhiệm vụ của đề tài** 7](#_Toc76762880)

[**3.** **Phương pháp nghiên cứu** 7](#_Toc76762881)

[**4.** **Đối tượng và phạm vi nghiên cứu** 7](#_Toc76762882)

[**CHƯƠNG 1: TỔNG QUAN VỀ CÔNG NGHỆ VÀ** 8](#_Toc76762883)

[**CÔNG CỤ SỬ DỤNG** 8](#_Toc76762884)

[**1.** **Tổng quan về công nghệ** 8](#_Toc76762885)

[**2.** **Công cụ sử dụng** 11](#_Toc76762886)

# **DANH MỤC HÌNH ẢNH**

# **DANH MỤC CÁC TỪ VIẾT TẮT**

# **LỜI MỞ ĐẦU**

1. **Mục tiêu nghiên cứu đề tài**

* Những khó khăn của doanh nghiệp trong việc quản lý nguồn dữ liệu lớn, là sự trộn lẫn các dữ liệu không liên quan tới nhau. Việc này có thể làm doanh nghiệp mất thời gian trong việc sắp xếp các dữ liệu, dẫn đến việc mất các dữ liệu quan trọng. Nhưng khi các dữ liệu được trình bày trên một Dashboard, thì những dữ liệu cần thiết sẽ liên quan mật thiết, chặt chẽ với nhau. Khách hàng dễ dàng phân tích dữ liệu cụ thể một cách chuyên sâu hơn. Dashboard ngoài việc dễ hiểu nhờ sự sắp xếp hợp lý, thì còn giúp người dùng dễ dàng tập trung nhiều hơn vào những chỉ số liên quan đến các chiến dịch đề ra
* Xuất phát từ những vấn đề trên, em được giao đề tài thực hiện “TRỰC QUAN HÓA DỮ LIỆU CHO KHÁCH HÀNG”.

1. **Nhiệm vụ của đề tài**

* Nghiên cứu ngôn ngữ Python đặc biệt là thư viện dash, pandas, plotly – các thư viện hỗ trợ cho việc xây dựng Dashboard.
* Nghiên cứu và tạo các thủ tục nâng cao hỗ trợ cho việc tương tác các Dropdownlist, biểu đồ, bảng.
* Nghiên cứu và xây dựng Dashboard bằng ngôn ngữ Python để phục vụ khách hàng.

1. **Phương pháp nghiên cứu**

* Phương pháp thu thập dữ liệu.
* Phương pháp phân tích dữ liệu.

1. **Đối tượng và phạm vi nghiên cứu**

* Để đạt được mục tiêu của đề tài, nhóm tập trung nghiên cứu các vấn đề sau:
* Về lý thuyết
* Tìm hiểu Dashboard
* Tìm hiểu phân hệ tài chính
* Tìm hiểu ngôn ngữ lập trình Python
* Tìm hiểu cách viết thủ tục nâng cao trong SQL Server
* Tìm hiểu dữ liệu của khách hàng
* Về kỹ thuật
* Đọc hiểu nghiệp vụ của khách hàng
* Tạo các thủ tục nâng cao trong SQL Server
* Xây dựng Dashboard dựa vào ngôn ngữ Python (package dash, pandas, pymsql,...)

**CHƯƠNG 1:**

**TỔNG QUAN VỀ CÔNG NGHỆ VÀ** **CÔNG CỤ SỬ DỤNG**

1. **Tổng quan về công nghệ**

* **DASH**

1. Dash của Python được viết trên Flask, Plotly.js và React.js, Dash lý tưởng để xây dựng các ứng dụng trực quan hóa dữ liệu với giao diện người dùng tùy chỉnh cao bằng Python thuần túy, phù hợp làm việc với dữ liệu bằng Python.

* Dash là mã nguồn mở và các ứng dụng của nó chạy trên trình duyệt web.
* Mục tiêu là tạo trang tổng quan ứng dụng web với các kỹ năng phát triển web tối thiểu hoặc không có như Javascript. Mặc dù Dash Python cho phép tùy chỉnh tốt hơn với việc sử dụng ngăn xếp Javascript, HTML, CSS
* Dash đi kèm với thư viện các thành phần cốt lõi, cho phép dễ dàng khai báo các thành phần HTML phổ biến như Dropdowns hoặc Graphs.
* Một điều thuận tiện khi Plotly graph\_objects được kết nối với Dash là bạn tự động nhận được các tính năng của Plotly mà không cần thêm mã. Điều này bao gồm thu phóng, tính năng di chuột, so sánh di chuột, tùy chọn tải xuống, chụp ảnh,…

*Bố cục ứng dụng dash*

* Một ứng dụng Dash thường bao gồm hai phần. Phần đầu tiên là bố cục và mô tả ứng dụng và phần thứ hai là mô tả khả năng tương tác của ứng dụng.
* ***dash\_html\_components:*** cung cấp các lớp HTML cho phép tạo nội dung HTML bằng Python. Ngoài ra, có thể tạo các thành phần tùy chỉnh của riêng mình bằng cách sử dụng Javascript và React Js.
* ***dash\_core\_components:*** tạo một danh sách thả xuống, chuyển các tùy chọn dưới dạng danh sách các từ điển, các nút radio sử dụng RadioItems, gọi Checklist tạo các hộp kiểm tra và Input – văn bản đầu vào trong ứng dụng.
* Để có thể tương tác trong ứng dụng Dash, chúng ta cần nhập Input và Output từ dash.dependencies. Để kích hoạt, Dash này cung cấp một trình trang trí @app để có thể liên kết một hạn gọi lại id đã tạo và trường đầu vào HTML.
* **PANDAS**
* Thư viện pandas trong Python là một thư viện mã nguồn mở, hỗ trợ đắc lực trong thao tác dữ liệu. Đây cũng là một công cụ phân tích và xử lý dữ liệu mạnh mẽ của ngôn ngữ lập trình Python.
* Thư viện này sử dụng một cấu trúc dữ liệu riêng là DataFrame. Đồng thời, pandas cũng cung cấp rất nhiều chức năng xử lý và làm việc trên cấu trúc dữ liệu này.
* cho phép đọc/ ghi dữ liệu giữa bộ nhớ và nhiều định dạng file: csv, text, excel, sql,…
* Liên kết dữ liệu thông minh, xử lý được trường hợp dữ liệu bị thiếu. Tự động đưa dữ liệu lộn xộn về dạng có cấu trúc.
* Dễ dàng thay đổi bố cục của dữ liệu.
* Tích hợp cơ chế trượt, lập chỉ mục, lấy ra tập con từ tập dữ liệu lớn.
* Có thể thêm, xóa các dữ liệu.
* Tập hợp hoặc thay đổi dữ liệu với group by cho phép thực hiện các toán tử trên tập dữ liệu.
* Hiệu quả cao trong kết hợp các tập dữ liệu.
* **PLOTLY**
* Plotly là một thư viện hướng mã nguồn mở, hỗ trợ người dùng trực quan hóa dữ liệu tương tác. Đồng thời, thư viện này cho phép vẽ đồ thị tương tác không những trực quan hơn mà còn thuyết phục người nghe hơn.
* Hỗ trợ hơn 40 loại bảng biểu khác nhau cho rất nhiều ứng dụng về thống kê, tài chính, địa lý, khoa học và các ứng dụng đời thực khác.
* Cho phép người dùng Python tạo ra các thiết kế web trực quan, sinh động để biểu thị trên nhiều loại cửa sổ như Visual Studio Code, Jupyter.
* Lưu trữ trong tập HTML độc lập hoặc tích hợp trong các ứng dụng web thuần Python thông qua Dash.

1. **Công cụ sử dụng**

* Một số công cụ phổ biến dùng để Viết plotly Dash : Visua studio code, sublime text3, pycharm, visua studio…
* Công cụ đi kèm: SQL SERVER, mysql