

PHÂN TÍCH DỮ LIỆU TUYỂN DỤNG TRONG LĨNH VỰC CÔNG NGHỆ THÔNG TIN

DỰ ÁN TUYỂN DỤNG

Sinh viên thực hiện: Đỗ Quang Hùng
MSSV: 20185365
Giảng viên hướng dẫn: ThS.Nguyễn Danh Tú

Hà Nội, tháng 8 năm 2023



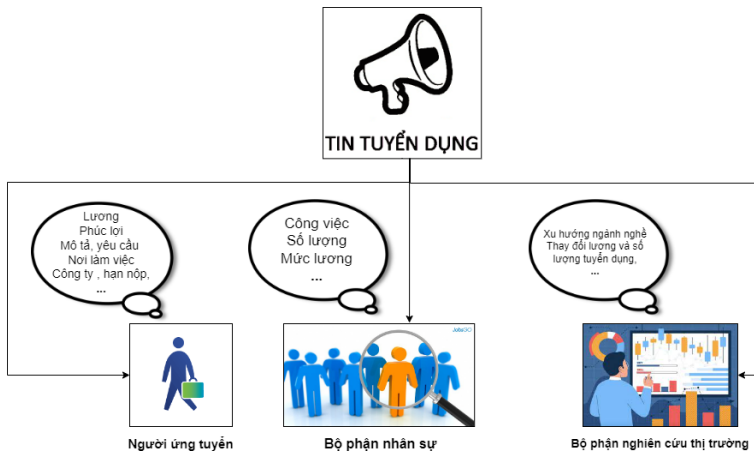
Nội dung

1 Khảo sát

2 Phân tích và thiết kế

3 Xây dựng chương trình

Nhu cầu xây dựng hệ thống



Quy trình xử lý tổng quát



Thu thập dữ liệu

Việc làm IT

17 Data Extracted

Task paused

Duplicates: 0 Item(s)

Time Spent: 1m 12s

Avg. Speed: 14 lines/min

Octoparse

Resume Stop

Task Overview Data List Event Log Recent Runs

#	Text	Text1	Text2	Title	Title_URL	Salary	Company	Address	Title1	File
9	Chuyên Viên IT Hệ ...	Thỏa thuận	HA Ngi	System Administra...	https://news.bipon...	20 - 40 triệu	Công ty cổ phần A...	HA Ngi	System Administra...	
10	Chuyên Viên IT Hệ ...	Thỏa thuận	HA Ngi	Nhân Viên Giám S...	https://news.bipon...	6 - 10 triệu	CÔNG TY CỔ PHẦN...	HA Ngi	Nhân Viên Giám S...	
11	Chuyên Viên IT Hệ ...	Thỏa thuận	HA Ngi	Control Tàng Nhậ...	https://news.bipon...	Từ 22 triệu	Công ty Cổ phần Y...	HA Ngi	Control Tàng Nhậ...	
12	Chuyên Viên IT Hệ ...	Thỏa thuận	HA Ngi	Product Owner	https://news.bipon...	Thỏa thuận	Công ty Cổ phần ...	HA Ngi	Product Owner	
13	Chuyên Viên IT Hệ ...	Thỏa thuận	HA Ngi	Solution Architect (...)	https://news.bipon...	Từ 3.000 USD	Công ty CP Kinh đ...	HA Ngi	Solution Architect (...)	
14	Chuyên Viên IT Hệ ...	Thỏa thuận	HA Ngi	Front-End Engineer...	https://news.bipon...	Từ 1.000 USD	CÔNG TY TNHH S...	Hà Chí Minh	Front-End Engineer...	
15	Chuyên Viên IT Hệ ...	Thỏa thuận	HA Ngi	Digital Product Ow...	https://news.bipon...	Thỏa thuận	Ngân Hàng Thươn...	HA Ngi	Digital Product Ow...	
16	Chuyên Viên IT Hệ ...	Thỏa thuận	HA Ngi	IT Communicator	https://news.bipon...	Từ 1.500 USD	Công ty TNHH Hậ...	HA Ngi	IT Communicator	
17	Chuyên Viên IT Hệ ...	Thỏa thuận	HA Ngi	Android Developer...	https://news.bipon...	Từ 45 triệu	Humax Vietnam	HA Ngi	Android Developer...	

< 1 > Go to Page

Export local data (Việc làm IT)

Export data as

- ☒ Excel
 ☐ CSV
 ☐ HTML
 ☐ JSON
 ☐ Xml

Export to database

- ☐ Google Sheets
 ☐ SqlServer
 ☐ MySql

Dữ liệu đưa vào cơ sở dữ liệu

STT	Tên trường	Mô tả	Kiểu dữ liệu
1	ID	Số định danh duy nhất cho mỗi bản ghi	varchar(20)
2	Web	Trang web đăng tin tuyển dụng	varchar(20)
3	TenCV	Tên vị trí tuyển dụng	text
4	CongTy	Tên công ty tuyển dụng	text
5	DiaDiem	Địa điểm làm việc	varchar(100)
6	Luong	Mức lương tuyển dụng	varchar(100)
7	LoaiHinh	Hình thức làm việc	varchar(50)
8	KinhNghiem	Kinh nghiệm làm việc	varchar(100)
9	CapBac	Cấp bậc trong công ty	varchar(50)
10	HanNop	Hạn nộp hồ sơ ứng tuyển	varchar(20)
11	YeuCau	Yêu cầu công việc	text
12	SoLuong	Số lượng tuyển cho vị trí tuyển dụng	int

Nội dung

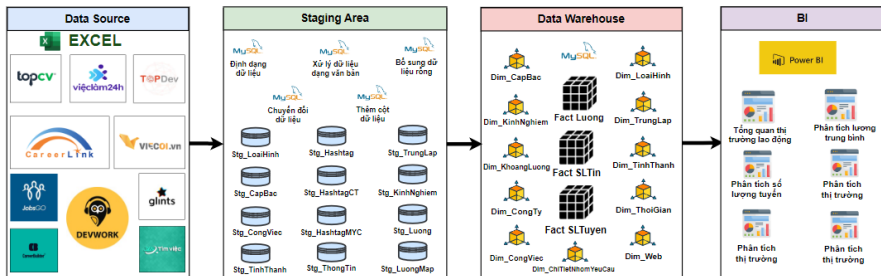
1 Khảo sát

2 **Phân tích và thiết kế**

3 Xây dựng chương trình

Thiết kế kiến trúc tổng thể

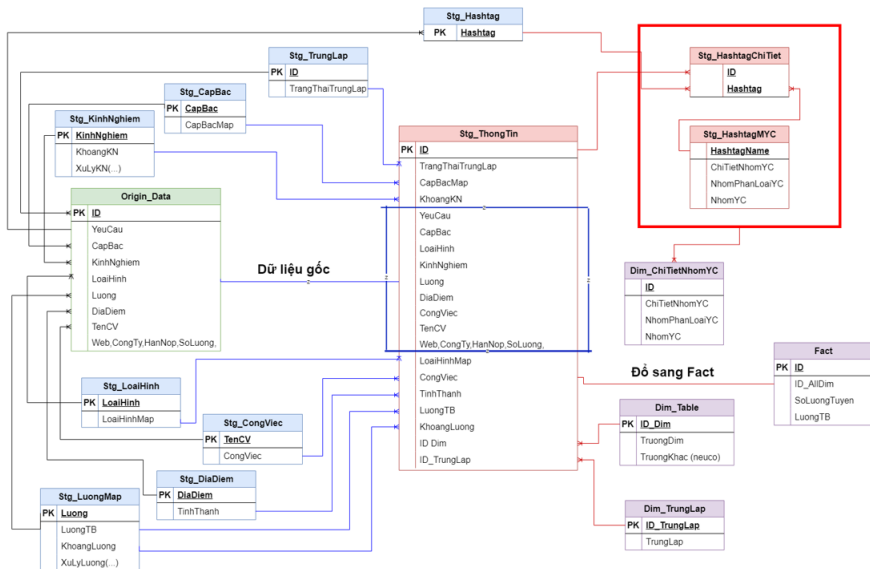
Kiến trúc tổng thể toàn bộ quá trình từ thu thập đến xử lý, lưu trữ và sử dụng dữ liệu.



Phân tích cấu trúc tin tuyển dụng

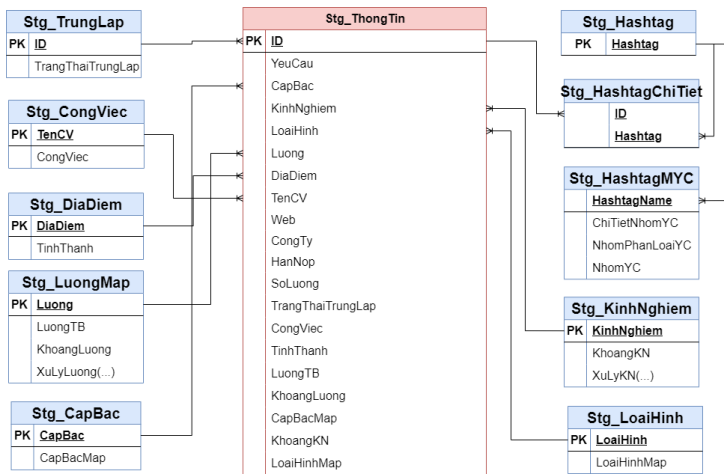
Trang	Địa điểm	Loại hình	Cấp bậc	Kinh Nghiệm	Lương	Số lượng
CareerBuilder	Tỉnh thành VD: Hà Nội	Toàn thời gian Bán thời gian Hình thức khác Thực tập	Giám đốc Nhân viên Mới tốt nghiệp Quản lý Trưởng nhóm / Giám sát	Năm	VND USD Cạnh tranh	Thu thấp tại web Nếu web không có thì mặc định = 1
	Tỉnh thành VD: Hà Nội	Toàn thời gian Nhân viên hợp đồng Nhân viên toàn thời gian Nhân viên toàn thời gian tạm thời	Nhân viên toàn thời gian Nhân viên Sinh viên / Thực tập sinh Mới đi làm Kỹ thuật viên / Kỹ sư Trưởng nhóm / Giám sát Quản lý / Trưởng phòng Nhân viên toàn thời gian tạm thời	Năm	VND USD Thương lượng Cạnh tranh	Thu thấp tại web Nếu web không chứa giá trị thì mặc định số lượng tuyển là 1
Vieclam24h	Tỉnh thành VD: Hà Nội Nhiều tỉnh thành VD: Hà Nội, Hồ Chí Minh , Đà Nẵng	Toàn thời gian cố định Bán thời gian cố định Thực tập Theo hợp đồng tư vấn Bán thời gian tạm thời Toàn thời gian tạm thời Khác	Chuyên viên- nhân viên Chuyên gia Cộng tác viên Quản lý cấp cao Quản lý cấp trung Quản lý nhóm- giám sát	Năm Chưa có kinh nghiệm	VND Thỏa thuận	Thu thấp tại web Nếu web không chứa giá trị thì mặc định số lượng tuyển là 1
	Tỉnh thành VD: Hà Nội Nhiều tỉnh thành	Fulltime In Office Hybrid	Junior Senior Middle All Levels Fresher	Không yêu cầu kinh nghiệm Tất cả kinh nghiệm thấp	Thương lượng USD	Thu thấp tại web Nếu web không chứa giá trị thì mặc định số lượng tuyển là 1
TopDev	Tỉnh thành VD: Hà Nội Nhiều tỉnh thành VD: Hà Nội, Hồ Chí Minh , Đà Nẵng	Fulltime In Office Hybrid Remote	Junior Senior Middle All Levels Fresher Trưởng Nhóm Trưởng phòng Nhân viên	Không yêu cầu kinh nghiệm Tất cả kinh nghiệm thấp năm	Thương lượng USD VND	Thu thấp tại web Nếu web không chứa giá trị thì mặc định số lượng tuyển là 1
DevWork	Tỉnh thành VD: Hà Nội	Toàn thời gian Bán thời gian	Middle Senior PM - Project Manager Junior Fresher Team Leader Super Manager Internship	Không yêu cầu kinh nghiệm Năm	Cạnh tranh VND	Thu thấp tại web Nếu web không chứa giá trị thì mặc định số lượng tuyển là 1
	Tỉnh thành VD: Hà Nội Nhiều tỉnh thành VD: Hà Nội, Hồ Chí Minh, Đà Nẵng	Full-time Part-time	Nhân Viên/Chuyên Viên Trưởng Nhóm/Trưởng Phòng Thực Tập Sinh Giám Đốc Và Cấp Cao Hơn	Không yêu cầu Năm	Thỏa thuận VND	Thu thấp tại web Nếu web không chứa giá trị thì mặc định số lượng tuyển là 1
TimVec365	Tỉnh thành VD: Hà Nội Nhiều tỉnh thành VD: Hà Nội & 2 nơi khác	Toàn thời gian cố định khác Toàn thời gian Bán thời gian Toàn thời gian tạm thời Hợp đồng	Nhân viên Thực tập sinh Mới tốt nghiệp Trưởng nhóm Phó trưởng phòng Trưởng phòng Giám đốc Tổng giám đốc Quản lý cấp cao Quản lý cấp trung	Không yêu cầu kinh nghiệm Năm	Thỏa Thuận VND USD	Thu thấp tại web Nếu web không chứa giá trị thì mặc định số lượng tuyển là 1

Thiết kế mô hình dữ liệu toàn hệ thống



Thiết kế vùng đệm

Mô hình dữ liệu logic vùng đệm:



SELECT * FROM Pgm_Tungtung; --						
ID	envCmp	envComp	envDateIn	envDateOut	ID_Tungtung	
1	Net	Net	Net	Net	Net	Net
2	Net	Net	Net	Net	Net	Net
3	Net	Net	Net	Net	Net	Net
4	Net	Net	Net	Net	Net	Net
5	Net	Net	Net	Net	Net	Net
6	Net	Net	Net	Net	Net	Net
7	Net	Net	Net	Net	Net	Net
8	Net	Net	Net	Net	Net	Net
9	Net	Net	Net	Net	Net	Net
10	Net	Net	Net	Net	Net	Net
11	Net	Net	Net	Net	Net	Net
12	Net	Net	Net	Net	Net	Net
13	Net	Net	Net	Net	Net	Net
14	Net	Net	Net	Net	Net	Net
15	Net	Net	Net	Net	Net	Net
16	Net	Net	Net	Net	Net	Net
17	Net	Net	Net	Net	Net	Net
18	Net	Net	Net	Net	Net	Net
19	Net	Net	Net	Net	Net	Net
20	Net	Net	Net	Net	Net	Net
21	Net	Net	Net	Net	Net	Net
22	Net	Net	Net	Net	Net	Net
23	Net	Net	Net	Net	Net	Net
24	Net	Net	Net	Net	Net	Net
25	Net	Net	Net	Net	Net	Net
26	Net	Net	Net	Net	Net	Net
27	Net	Net	Net	Net	Net	Net
28	Net	Net	Net	Net	Net	Net
29	Net	Net	Net	Net	Net	Net
30	Net	Net	Net	Net	Net	Net
31	Net	Net	Net	Net	Net	Net
32	Net	Net	Net	Net	Net	Net
33	Net	Net	Net	Net	Net	Net
34	Net	Net	Net	Net	Net	Net
35	Net	Net	Net	Net	Net	Net
36	Net	Net	Net	Net	Net	Net
37	Net	Net	Net	Net	Net	Net
38	Net	Net	Net	Net	Net	Net
39	Net	Net	Net	Net	Net	Net
40	Net	Net	Net	Net	Net	Net
41	Net	Net	Net	Net	Net	Net
42	Net	Net	Net	Net	Net	Net
43	Net	Net	Net	Net	Net	Net
44	Net	Net	Net	Net	Net	Net
45	Net	Net	Net	Net	Net	Net
46	Net	Net	Net	Net	Net	Net
47	Net	Net	Net	Net	Net	Net
48	Net	Net	Net	Net	Net	Net
49	Net	Net	Net	Net	Net	Net
50	Net	Net	Net	Net	Net	Net
51	Net	Net	Net	Net	Net	Net
52	Net	Net	Net	Net	Net	Net
53	Net	Net	Net	Net	Net	Net
54	Net	Net	Net	Net	Net	Net
55	Net	Net	Net	Net	Net	Net
56	Net	Net	Net	Net	Net	Net
57	Net	Net	Net	Net	Net	Net
58	Net	Net	Net	Net	Net	Net
59	Net	Net	Net	Net	Net	Net
60	Net	Net	Net	Net	Net	Net
61	Net	Net	Net	Net	Net	Net
62	Net	Net	Net	Net	Net	Net
63	Net	Net	Net	Net	Net	Net
64	Net	Net	Net	Net	Net	Net
65	Net	Net	Net	Net	Net	Net
66	Net	Net	Net	Net	Net	Net
67	Net	Net	Net	Net	Net	Net
68	Net	Net	Net	Net	Net	Net
69	Net	Net	Net	Net	Net	Net
70	Net	Net	Net	Net	Net	Net
71						

Stg_LuongMap 1 x

SELECT * FROM Stg_LuongMap of Enter a SQL expression for filter results (use Ctrl+Space)

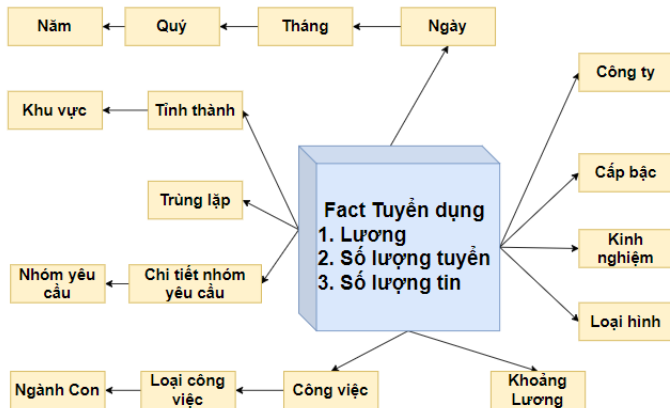
Grid	Luong	DonVi	MinLuong	MaxLuong	QuyDoiMaxLuong	QuyDoiMaxLuong	LuongTB	ChaoLuong	
1	lưu trữ lương	trình lương	0	1	0	0	0	0	Thùng lương
2	8 - 9 tr VND	trình VND	8	9	8	9	8.5	3 - 10 tr	
3	10 tr - 13 tr VND	trình VND	10	13	10	13	11.5	10 - 15 tr	
4	15 tr - 20 tr VND	trình VND	15	20	15	20	17.5	15 - 25 tr	
5	8 - 13 tr VND	trình VND	8	13	8	13	10.5	10 - 15 tr	
6	12 tr - 15 tr VND	trình VND	12	15	12	15	13.5	10 - 15 tr	
7	20 tr - 30 tr VND	trình VND	20	30	20	30	25	20 - 30 tr	
8	10 tr - 15 tr VND	trình VND	10	15	10	15	12.5	10 - 15 tr	
9	8 tr - 16 tr VND	trình VND	8	16	8	16	12	10 - 15 tr	
10	trình 8 tr VND	trình VND	8	8	8	8	8	3 - 10 tr	
11	17 tr - 35 tr VND	trình VND	17	35	17	35	26	25 - 35 tr	
12	13 tr - 20 tr VND	trình VND	13	20	13	20	16.5	15 - 25 tr	
13	45 tr - 50 tr VND	trình VND	45	50	45	50	47.5	35 - 50 tr	
14	20 tr - 25 tr VND	trình VND	20	25	20	25	22.5	15 - 25 tr	
15	12 tr - 20 tr VND	trình VND	12	20	12	20	16	15 - 20 tr	
16	13 tr - 15.5 tr VND	trình VND	13	15.5	13	15.5	14.3	10 - 15 tr	
17	8 tr - 11 tr VND	trình VND	8	11	8	11	9.5	3 - 10 tr	
18	500 - 1,000 USD		500	1,000	11.8	23.5	17.7	15 - 25 tr	
19	lên đến 1,000 USD		1,000	1,000	23.5	23.5	23.5	15 - 25 tr	
20	lên đến 7.5 tr VND		7.5	7.5	7.5	7.5	7.5	3 - 10 tr	
21	15 tr - 25 tr VND		15	25	15	25	20	15 - 25 tr	
22	600 - 800 USD		600	800	14.1	18.8	16.5	15 - 25 tr	
23	2,000 - 3,000 USD		2,000	3,000	47	70.5	58.8	Tên 50 tr	
24	lên đến 12 tr VND		12	12	12	12	12	10 - 15 tr	

Phân tích danh mục chiều khái niệm

Trùng lặp	Cấp bậc	Kinh nghiệm	Khoảng lương	Khu vực	Loại hình	Nhóm yêu cầu	Ngành con	Thời gian
Có	Thực tập sinh	Không yêu cầu kinh nghiệm	Dưới 3 triệu	Bắc	Bán thời gian	Kiến thức	Quản lý	
Không	Chuyên viên	Dưới 1 năm	3-10 triệu	Trung	Toàn thời gian	Kỹ năng	IT Support	Ngày
	Trưởng nhóm	1-3 năm	10-15 triệu	Nam	Hình thức khác	Công cụ	Project Manager	
	Quản lý	3-5 năm	15-25 triệu	Khác		Chứng chỉ	Quản trị hệ thống	Tháng
	Trưởng/ Phó phòng	Trên 5 năm	25-35 triệu				Thiết kế	
	Giám đốc/ Phó GD		35-50 triệu		Công ty	Chi tiết nhóm yêu cầu	Lập trình	Quý
			Hơn 50 triệu				Data	
			Thỏa thuận		Web	Công việc	Khác	Năm

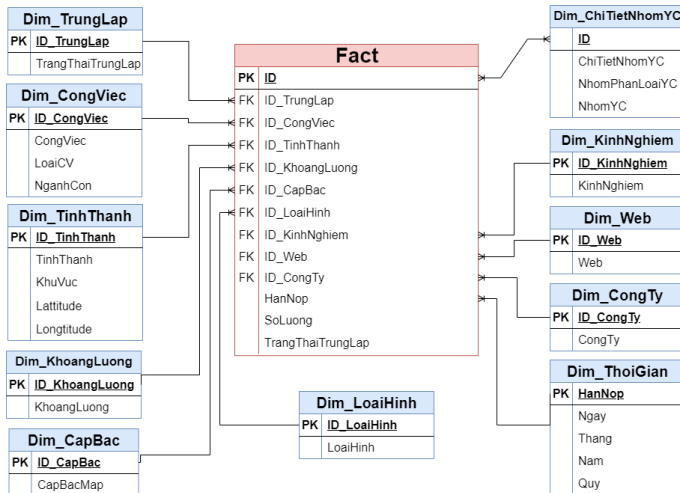
Thiết kế kho chủ đề

Mô hình dữ liệu logic của kho chủ đề:



Thiết kế kho chủ đề

Mô hình dữ liệu vật lý của kho chủ đề:



Nội dung

- 1 Khảo sát
- 2 Phân tích và thiết kế
- 3 Xây dựng chương trình**

Xây dựng chương trình

Mục đích	Mã	Roles	Mã	Chủ điểm	Mã	Chu kỳ trên đồ thị	Mã
Phân tích	P	Người lao động	LĐ	Tổng quan	1xx	Không có thời gian	x0x
Tổng hợp	T	Bộ phận Nhân sự	NS	Số lượng tuyển	2xx	Năm	x1x
		Nghiên cứu thị trường	NC	Lương	3xx		
				Số lượng tin	4xx		

Danh sách các chiều phân tích

NgànhCon
☐ Data
 ☐ IT Support
 ☐ Khác
 ☐ Lập trình
 ☐ Project Manager
 ☐ Quản lý
 ☐ Quản trị hệ thống
 ☐ Thiết kế

KhoangLương
☐ 10 - 15 triệu
 ☐ 15 - 25 triệu
 ☐ 25 - 35 triệu
 ☐ 3 - 10 triệu
 ☐ 35 - 50 triệu
 ☐ Dưới 3 triệu
 ☐ Thương lượng
 ☐ Trên 50 triệu

Nam
☐ 2022
 ☐ 2023
 ☐ 2024

Quy
☐ 1
 ☐ 2
 ☐ 3
 ☐ 4

Ngay
☐ 1
 ☐ 2
 ☐ 3
 ☐ 4
 ☐ 5
 ☐ 6
 ☐ 7
 ☐ 8
 ☐ 9
 ☐ 10
 ☐ 11
 ☐ 12
 ☐ 13
 ☐ 14
 ☐ 15
 ☐ 16
 ☐ 17
 ☐ 18
 ☐ 19
 ☐ 20
 ☐ 21
 ☐ 22
 ☐ 23
 ☐ 24
 ☐ 25
 ☐ 26
 ☐ 27
 ☐ 28
 ☐ 29
 ☐ 30
 ☐ 31

NhomYeuCau
☐ Chứng chỉ
 ☐ Công cụ
 ☐ Kiến thức
 ☐ Kỹ năng

CapBac
☐ Giám đốc/Phó GD
 ☐ Nhân viên
 ☐ Quản lý
 ☐ Thực tập sinh
 ☐ Trưởng nhóm
 ☐ Trưởng/Phó phòng

KhuVuc
☐ Bắc
 ☐ Khác
 ☐ Nam
 ☐ Trung

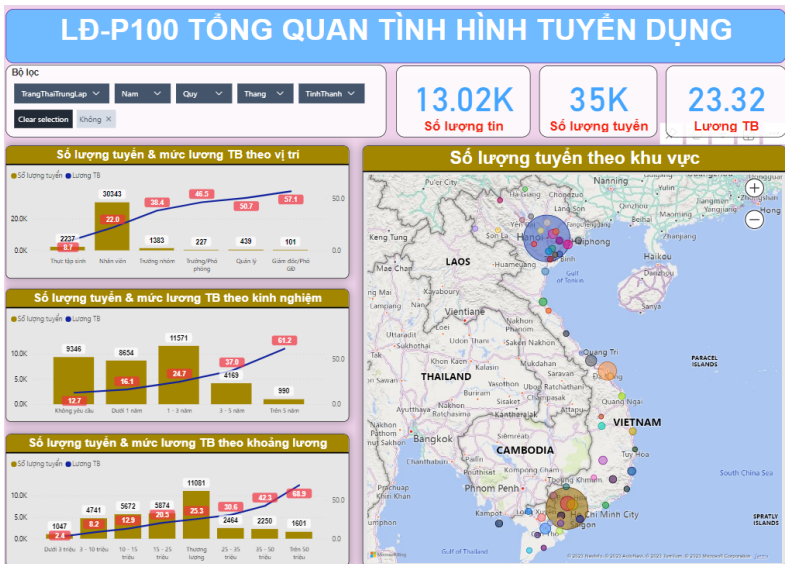
TrungLap
☐ Có
 ☐ Không

TinhThanh
☐ An Giang
 ☐ Bà Rịa - Vũng Tàu
 ☐ Bắc Giang
 ☐ Bạc Liêu
 ☐ Bắc Ninh
 ☐ Bến Tre
 ☐ Bình Định
 ☐ Bình Dương
 ☐ Bình Phước
 ☐ Bình Thuận
 ☐ Cà Mau
 ☐ Cần Thơ
 ☐ Đà Nẵng
 ☐ Đắk Lắk
 ☐ Đắk Nông
 ☐ Điện Biên
 ☐ Đồng Nai
 ☐ Đồng Tháp
 ☐ Gia Lai
 ☐ Hà Giang
 ☐ Hà Nam
 ☐ Hà Nội
 ☐ Hà Tĩnh
 ☐ Hải Dương
 ☐ Hải Phòng
 ☐ Hậu Giang
 ☐ Hồ Chí Minh

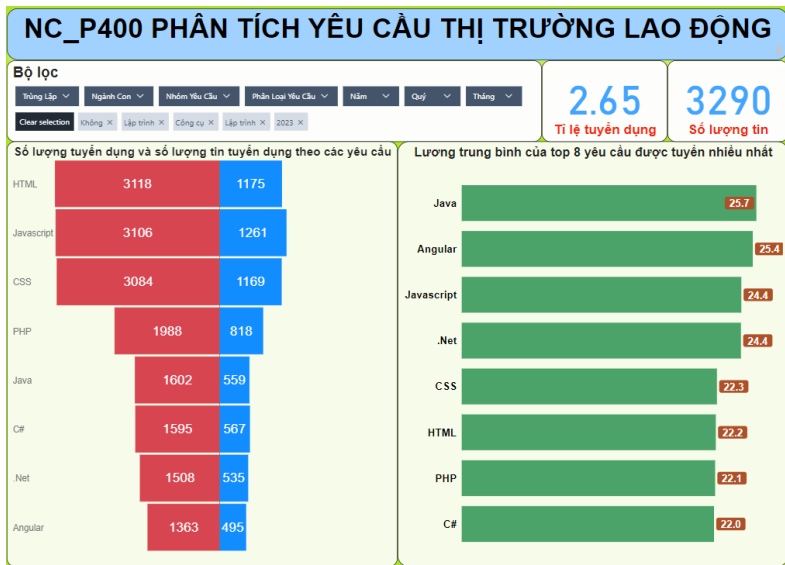
NhomPhanLoaiYeuCau
☐ Bằng cấp
 ☐ CCNA/CCNP/PCAP
 ☐ Có kinh nghiệm làm việc
 ☐ Data
 ☐ Đọc hiểu, giao tiếp ngoại ngữ
 ☐ Kiểm soát chất lượng
 ☐ Kỹ năng mềm
 ☐ Lập trình
 ☐ Mạng xã hội
 ☐ Marketing
 ☐ Microsoft Office
 ☐ MOS
 ☐ Ngoại ngữ
 ☐ Phần tích
 ☐ Quản trị
 ☐ Quảng cáo
 ☐ Thiết kế
 ☐ Thương mại điện tử

CongTy
☐ [The Coffee House] Công Ty TNHH T.
 ☐ 10In
 ☐ 1990 Agency
 ☐ 1OFFICE
 ☐ 2NF Software
 ☐ 3S Cloud Render Farm
 ☐ 3S Intersoft JSC

Xây dựng dashboard

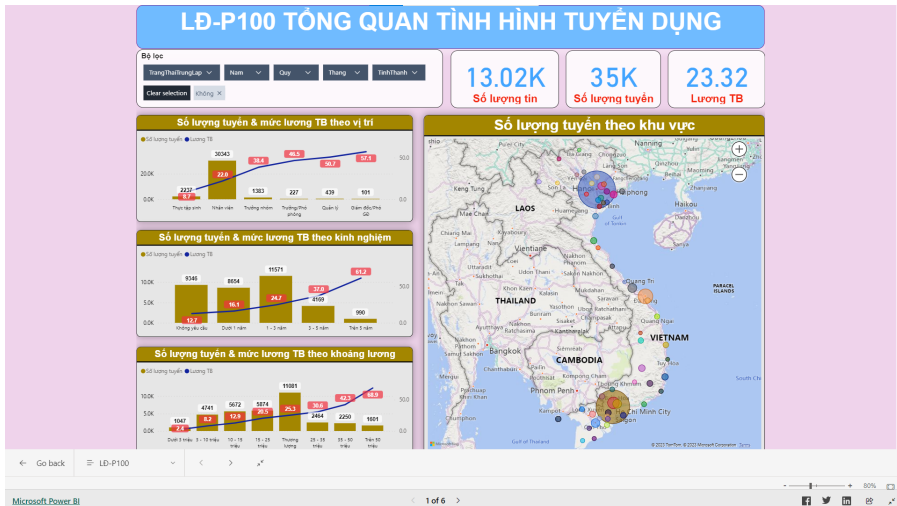


Xây dựng dashboard



Đưa dashboard lên Internet

Giao diện Dashboard khi được đưa lên internet:



Kết quả đạt được và hướng đề tài trong tương lai

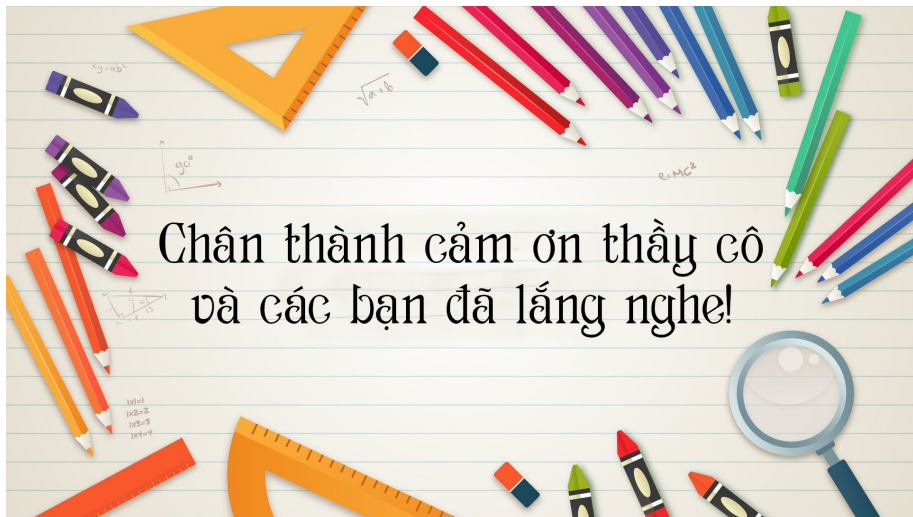
● Kết quả đạt được:

- Thu thập dữ liệu từ nhiều nguồn website khác nhau, xử lý dữ liệu thô và đưa vào cơ sở dữ liệu.
- Xây dựng được hệ thống các thủ tục để xử lý dữ liệu tự động cho cơ sở dữ liệu tuyển dụng.
- Tự động hóa quy trình cập nhật dữ liệu trên cơ sở dữ liệu từ khâu xử lý dữ liệu, đưa dữ liệu vào kho dữ liệu đến các báo cáo phân tích.
- Trực quan hóa dữ liệu bằng các báo cáo phân tích tổng quan về lĩnh vực công nghệ thông tin.

● Hướng đề tài trong tương lai:

- Tối ưu hoá hệ thống thu thập dữ liệu từ các website tuyển dụng.
- Cập nhật dữ liệu định kì hàng tháng để tăng kích thước bộ dữ liệu.
- Xây dựng thêm các báo cáo phân tích chuyên sâu.

Lời cảm ơn



Chân thành cảm ơn thầy cô
và các bạn đã lắng nghe!