



FPT POLYTECHNIC

BÁO CÁO KHO DỮ LIỆU

Chuyên ngành Xử Lý Dữ Liệu

Xây dựng kho dữ liệu tuyển dụng từ LinkedIn ngành CNTT

Lớp: DAT204.01

Nhóm: 4

Các thành viên:

1. Nguyễn Lan Anh - PH52788 (Nhóm trưởng)
2. Trần Phạm Quốc Khánh - PH51849
3. Nguyễn Thành Tâm - PH50613
4. Mai Hoàng Hà - PH53570
5. Lê Huy Hoàng - PH53268
6. Nguyễn Việt Anh - PH52136

GV HD: Chu Thị Ngân

Hà Nội - 2025

Mục lục

1. Giới thiệu.....	4
1.1. Giới thiệu dự án.....	4
1.2. Mục tiêu dự án.....	4
1.3. Ý nghĩa, ứng dụng của dự án.....	5
2. Tổng quan về kho dữ liệu.....	7
2.1. Kho dữ liệu (Data Warehouse).....	7
2.2. Giới thiệu về hệ thống kho dữ liệu nhóm phân tích.....	7
3. Mô tả hệ thống kho dữ liệu.....	8
4. Thiết kế CSDL.....	9
4.1. Xác định tập thực thể (liệt kê các thực thể tham gia trong hệ thống).....	9
4.2. Xác định các thuộc tính.....	9
4.3. Mô hình thực thể quan hệ.....	14
5. Xây dựng kho dữ liệu.....	15
5.1. Giới thiệu SQL Server Integration Services (SSIS).....	15
5.3. Chọn các chiều dữ liệu (Dimension).....	15
5.4. Xác định bảng sự kiện (Fact).....	16
5.5. Tích hợp dữ liệu và cài đặt kho dữ liệu.....	16
6. Khai thác kho dữ liệu.....	18
6.1. SQL Server Analysis Service (SSAS).....	18
6.2. SQL Server Analysis Service(SSAS) trong dự án.....	19
6.3. Khai thác dữ liệu từ kho dữ liệu.....	19
6.4. Phân tích và tổng hợp dữ liệu.....	20
6.5. Xây dựng báo cáo & trực quan hóa dữ liệu.....	22
6.6. Đưa ra kết luận phân tích.....	26
7. Tổng kết.....	27
8. Tài liệu tham khảo.....	31

Mục lục bảng

Bảng 1: fact_job_postings.....	10
Bảng 2: dim_companies.....	11
Bảng 3: dim_locations.....	12
Bảng 4: dim_industries.....	13
Bảng 5: dim_employment_types.....	13
Bảng 6: dim_seniority_levels.....	14
Bảng 7: dim_skills.....	14

Mục lục ảnh

Hình 1: Data Modeling.....	14
Hình 2: Mô hình lưu trữ.....	14
Hình 3: Luồng dữ liệu.....	17
Hình 4: Kết quả nạp dữ liệu vào SQL.....	18
Hình 5: Mô hình dữ liệu trên SQL Server Analysis Services.....	19
Hình 6: Mức lương trung bình theo các cấp bậc.....	20
Hình 7: Các kỹ năng được yêu cầu nhiều nhất.....	21
Hình 8: Mức lương giữa các ngành.....	21
Hình 9: Các công việc có nhiều ứng viên nhất.....	22
Hình 10: Công ty có số lượng công việc nhiều.....	22
Hình 11: Số lượng công việc theo các cấp bậc.....	23
Hình 12: Số lượng công việc theo hình thức tuyển dụng.....	23
Hình 13 : Kết quả trực quan hóa trong Power BI (Page Overview).....	24
Hình 14: Page Nhân sự.....	25
Hình 15: Page Công việc.....	26

1. Giới thiệu

1.1. Giới thiệu dự án

Trong thời đại số hóa, tuyển dụng nhân sự CNTT ngày càng thách thức, dù các nền tảng như LinkedIn giúp doanh nghiệp tiếp cận nhiều ứng viên hơn. Tuy nhiên, việc thu thập và phân tích dữ liệu tuyển dụng từ LinkedIn vẫn chưa được khai thác hiệu quả.

Dự án "Xây dựng kho dữ liệu tuyển dụng từ LinkedIn ngành CNTT" nhằm tạo kho dữ liệu tập trung, giúp doanh nghiệp phân tích xu hướng tuyển dụng, đánh giá ứng viên và tối ưu chiến lược nhân sự. Kho dữ liệu sẽ tích hợp thông tin về vị trí tuyển dụng, kỹ năng, mức lương, địa điểm làm việc và ứng viên tiềm năng.

1.2. Mục tiêu dự án

- Xây dựng kho dữ liệu tập trung về vị trí tuyển dụng, kỹ năng, mức lương, địa điểm làm việc.
- Phân tích xu hướng tuyển dụng, kỹ năng được yêu cầu nhiều nhất và mức lương theo ngành.
- Hỗ trợ doanh nghiệp đưa ra quyết định tuyển dụng hiệu quả và giúp ứng viên định hướng nghề nghiệp.

1.3. Ý nghĩa, ứng dụng của dự án

- Hỗ trợ tuyển dụng hiệu quả: Giúp doanh nghiệp nắm bắt xu hướng tuyển dụng, tối ưu quy trình tìm kiếm và đánh giá ứng viên.
- Định hướng phát triển kỹ năng: Ứng viên có thể biết được những kỹ năng nào đang được yêu cầu nhiều nhất để học tập và nâng cao năng lực.

- Dự báo thị trường lao động: Cung cấp dữ liệu phân tích về nhu cầu nhân lực, giúp doanh nghiệp và người lao động đưa ra quyết định đúng đắn.
- Thúc đẩy chuyển đổi số trong nhân sự: Ứng dụng dữ liệu lớn vào tuyển dụng giúp tối ưu chi phí và thời gian.

Dự án không chỉ hỗ trợ các doanh nghiệp trong quá trình tuyển dụng mà còn mang lại nhiều giá trị thiết thực cho ứng viên và các tổ chức giáo dục.

- Hỗ trợ tuyển dụng hiệu quả: Doanh nghiệp có thể nắm bắt xu hướng tuyển dụng trong ngành CNTT, tối ưu quy trình tìm kiếm và đánh giá ứng viên, đồng thời xác định được kỹ năng nào đang khan hiếm trên thị trường.
- Định hướng phát triển kỹ năng: Ứng viên có thể tra cứu những kỹ năng được yêu cầu nhiều nhất trong ngành, từ đó có kế hoạch học tập và rèn luyện phù hợp để nâng cao cơ hội việc làm.
- Dự báo thị trường lao động: Với việc phân tích dữ liệu từ nhiều nguồn, dự án cung cấp thông tin về sự thay đổi trong nhu cầu tuyển dụng, mức lương và các tiêu chí đánh giá ứng viên, giúp doanh nghiệp và người lao động đưa ra quyết định chính xác hơn.
- Thúc đẩy chuyển đổi số trong tuyển dụng: Việc ứng dụng dữ liệu lớn (Big Data) và kho dữ liệu (Data Warehouse) vào quá trình tuyển dụng giúp doanh nghiệp giảm thời gian và chi phí, đồng thời tăng hiệu quả tuyển dụng.
- Tối ưu hóa chiến lược nhân sự: Dựa vào dữ liệu phân tích, doanh nghiệp có thể đưa ra chiến lược tuyển dụng phù hợp hơn, xác định khu vực có nhiều nhân tài, điều chỉnh chính sách lương thưởng để thu hút ứng viên tốt nhất.

- Cải thiện chương trình đào tạo: Dữ liệu phân tích từ kho dữ liệu có thể giúp các trường đại học và tổ chức đào tạo hiểu rõ hơn về nhu cầu của thị trường, từ đó cập nhật chương trình giảng dạy để đáp ứng nhu cầu của doanh nghiệp.

Kết luận

Dự án "Xây dựng kho dữ liệu tuyển dụng từ LinkedIn ngành CNTT" là một bước tiến quan trọng trong việc ứng dụng dữ liệu lớn vào lĩnh vực tuyển dụng. Bằng cách tập trung và phân tích dữ liệu tuyển dụng từ LinkedIn, dự án giúp doanh nghiệp, ứng viên và tổ chức giáo dục có được thông tin giá trị, hỗ trợ ra quyết định chính xác và hiệu quả hơn. Trong bối cảnh ngành CNTT đang phát triển nhanh chóng, việc có một hệ thống kho dữ liệu tuyển dụng đáng tin cậy sẽ là chìa khóa giúp doanh nghiệp và người lao động tận dụng tối đa cơ hội trong thị trường việc làm.

2. Tổng quan về kho dữ liệu

2.1. Kho dữ liệu (Data Warehouse)

- Là một dạng kho dữ liệu giúp quản lý dữ liệu hệ thống cho phép hỗ trợ các hoạt động kinh doanh một cách thông minh.
- Chỉ nhằm mục đích thực hiện các truy vấn và phân tích, thường chứa một lượng lớn dữ liệu.
- Là cốt lõi của hệ thống BI được xây dựng để phân tích và báo cáo dữ liệu.

2.2. Giới thiệu về hệ thống kho dữ liệu nhóm phân tích

- ❖ Hệ thống kho dữ liệu tuyển dụng ngành CNTT thu thập, lưu trữ và phân tích dữ liệu từ LinkedIn, giúp đánh giá xu hướng tuyển dụng, yêu cầu kỹ năng và mức lương trung bình.
- ❖ Mục đích của hệ thống
 - Cung cấp dữ liệu tập trung về thị trường tuyển dụng CNTT.
 - Hỗ trợ doanh nghiệp xác định xu hướng tuyển dụng và mức lương.
 - Giúp ứng viên định hướng phát triển kỹ năng.
 - Hỗ trợ tổ chức giáo dục điều chỉnh chương trình đào tạo.
- ❖ Nguồn dữ liệu đầu vào
 - Thông tin tuyển dụng: vị trí, công ty, địa điểm, mức lương, kỹ năng.
 - Hồ sơ ứng viên: kinh nghiệm, kỹ năng, trình độ học vấn.
 - Dữ liệu từ báo cáo ngành và khảo sát lao động
- ❖ Cách tổ chức dữ liệu trong kho
 - Bảng sự kiện (Fact Table): Chứa thông tin bài đăng tuyển dụng, số lượng tuyển và mức lương trung bình.
 - Bảng chiều (Dimension Tables): Thông tin về công ty, vị trí, kỹ năng, khu vực và thời gian đăng tuyển.
- ❖ Lợi ích khi triển khai hệ thống
 - Cung cấp dữ liệu đáng tin cậy để phân tích xu hướng tuyển dụng.
 - Tối ưu hóa tuyển dụng với thông tin chi tiết về kỹ năng và lương.
 - Giúp doanh nghiệp định hình chiến lược nhân sự.
 - Hỗ trợ dự đoán sự phát triển của ngành CNTT dựa trên dữ liệu lịch sử.

3. Mô tả hệ thống kho dữ liệu

Hệ thống kho dữ liệu tuyển dụng ngành CNTT được xây dựng nhằm tổng hợp, lưu trữ và phân tích dữ liệu tuyển dụng từ LinkedIn, giúp các bên liên quan có cái nhìn toàn diện về thị trường lao động CNTT.

Vấn đề cần giải quyết:

- Phân tán dữ liệu: Dữ liệu tuyển dụng từ LinkedIn bị phân tán, khó khai thác và tổng hợp.
- Thiếu công cụ phân tích: Các doanh nghiệp và ứng viên gặp khó khăn trong việc đánh giá xu hướng tuyển dụng, mức lương và kỹ năng quan trọng.
- Thiếu thông tin chiến lược: Các tổ chức giáo dục và nhà tuyển dụng chưa có cơ sở dữ liệu đầy đủ để điều chỉnh chương trình đào tạo hoặc chiến lược nhân sự.

Giải pháp của hệ thống:

- Thu thập và chuẩn hóa dữ liệu tuyển dụng từ LinkedIn.
- Xây dựng kho dữ liệu với cấu trúc rõ ràng, hỗ trợ phân tích theo nhiều tiêu chí (vị trí, kỹ năng, mức lương, xu hướng theo thời gian).
- Cung cấp báo cáo và trực quan hóa dữ liệu giúp doanh nghiệp, ứng viên và các tổ chức giáo dục đưa ra quyết định chính xác hơn.

4. Thiết kế CSDL

4.1. Xác định tập thực thể (liệt kê các thực thể tham gia trong hệ thống)

Dựa trên dữ liệu, hệ thống bao gồm 5 thực thể chính:

- fact_job_postings (Bảng Fact - Bài đăng tuyển dụng)
- dim_companies (Công ty)
- dim_locations (Địa điểm công ty)
- dim_industries (Ngành công nghiệp)
- dim_employment_types (Loại hình công việc)
- dim_seniority_levels (Cấp bậc công việc)
- dim_skills (Kỹ năng)

4.2. Xác định các thuộc tính

- Thực thể fact_job_postings

Bảng 1: fact_job_postings

Thuộc tính	Mô Tả
job_id (PK)	ID duy nhất của công việc
company_id (FK)	ID duy nhất của công ty
location_id (FK)	ID duy nhất của địa điểm
industry_id (FK)	ID duy nhất của ngành
employment_type_id (FK)	ID của loại hình công việc
seniority_level_id (FK)	ID của cấp bậc lao động
posted_datetime	Ngày và giờ bài đăng được đăng tải
applicants	Số lượng ứng viên đã ứng tuyển

posted_datetime	Mức lương tối thiểu đề xuất cho công việc
min_salary	Lương tối thiểu
max_salary	Lương tối đa

- Thực thể dim_companies (Công ty)

Bảng 2: dim_companies

Thuộc tính	Mô Tả
company_id (PK)	ID duy nhất của công ty
company_name	Tên công ty
company_industry	Ngành nghề hoạt động
company_size	Quy mô công ty
company_region	Khu vực hoạt động chính của công ty
company_founded	Năm thành lập công ty
company_specialties	Các lĩnh vực chuyên môn mà công ty tập trung

- Thực thể dim_locations (Địa điểm công ty)

Bảng 3: dim_locations

Thuộc tính	Mô Tả
location_id (PK)	Mã định danh duy nhất của địa điểm
company_country	Quốc gia của công ty
company_locality	Thành phố hoặc địa phương của công ty
company_street	Địa chỉ chi tiết của công ty

- Thực thể dim_industries (Ngành công nghiệp)

Bảng 4: dim_industries

Thuộc tính	Mô tả
industry_id (PK)	Mã định danh của ngành công nghiệp
industry_name	Tên ngành công nghiệp

- Thực thể dim_employment_types (Loại hình công việc)

Bảng 5: dim_employment_types

Thuộc tính	Mô Tả
employment_type_id (PK)	Mã định danh loại hình công việc
employment_type_name	Tên loại hình công việc

- Thực thể dim_seniority_levels (Cấp bậc công việc)

Bảng 6: dim_seniority_levels

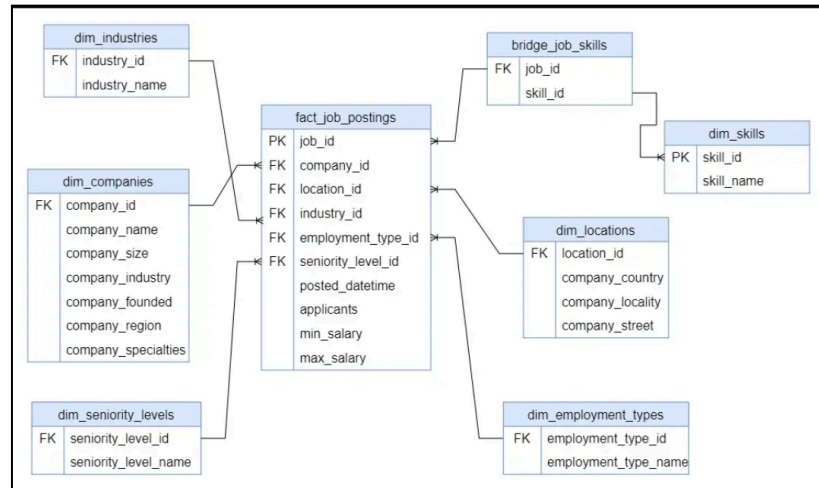
Thuộc tính	Mô tả
seniority_level_id (<i>PK</i>)	Mã định danh cấp bậc công việc
seniority_level_name	Tên cấp bậc công việc

- Thực thể dim_skills (Kỹ năng)

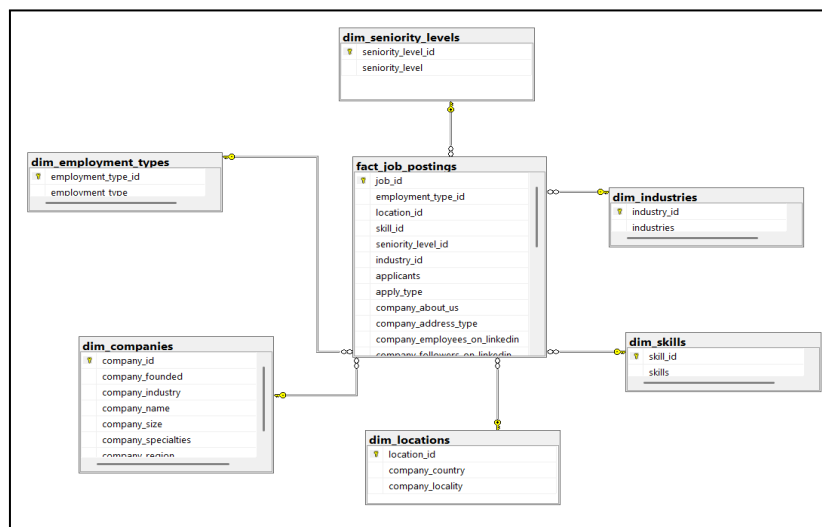
Bảng 7: dim_skills

Thuộc tính	Mô tả
skill_id (<i>PK</i>)	Mã định danh kỹ năng
job_id (<i>FK</i>)	Mã công việc
skill_name	Tên kỹ năng

4.3. Mô hình thực thể quan hệ



Hình 1: Data Modeling



Hình 2: Mô hình lưu trữ

5. Xây dựng kho dữ liệu

5.1. Giới thiệu SQL Server Integration Services (SSIS)

SQL Server Integration Services (SSIS) là một nền tảng tích hợp dữ liệu mạnh mẽ do Microsoft phát triển, được sử dụng để thực hiện các tác vụ **ETL (Extract, Transform, Load)** trong SQL Server. SSIS giúp doanh nghiệp xử lý và di chuyển dữ liệu từ nhiều nguồn khác nhau vào hệ thống đích một cách hiệu quả.

Dưới đây là các chức năng chính của SSIS:

- Trích xuất dữ liệu (Extract): Thu thập dữ liệu từ nhiều nguồn khác nhau như SQL Server, Oracle, Excel, CSV, API, JSON,...
- Biến đổi (Transform): Làm sạch dữ liệu, chuyển đổi kiểu dữ liệu, hợp nhất hoặc phân tách dữ liệu, tính toán giá trị mới.
- Tải dữ liệu (Load): Đưa dữ liệu đã xử lý vào cơ sở dữ liệu đích, kho dữ liệu (Data Warehouse) hoặc các hệ thống báo cáo.

5.2. Lợi ích của SQL Server Integration Services (SSIS)

- Tự động hóa - Tiết kiệm thời gian
- Xử lý dữ liệu lớn - Hiệu suất cao
- Tích hợp tốt với hệ sinh thái Microsoft
- Khả năng xử lý lỗi mạnh mẽ
- Dễ sử dụng và mở rộng

5.3. Chọn các chiều dữ liệu (Dimension)

- Dim Companies: Danh sách các công ty tuyển dụng.
- Dim Employment: Hình thức làm việc
- Dim Industries: Ngành nghề tuyển dụng
- Dim Locations: Vị trí địa lý công việc
- Dim Seniority Levels: Cấp bậc kinh nghiệm

- Dim Skills: Kỹ năng yêu cầu
- Fact job Postings: Dữ liệu tuyển dụng

5.4. Xác định bảng sự kiện (Fact)

- fact_job_postings: Bảng chính chứa dữ liệu tuyển dụng, bao gồm thông tin về công việc, công ty, địa điểm, lương, v.v.

5.5. Tích hợp dữ liệu và cài đặt kho dữ liệu

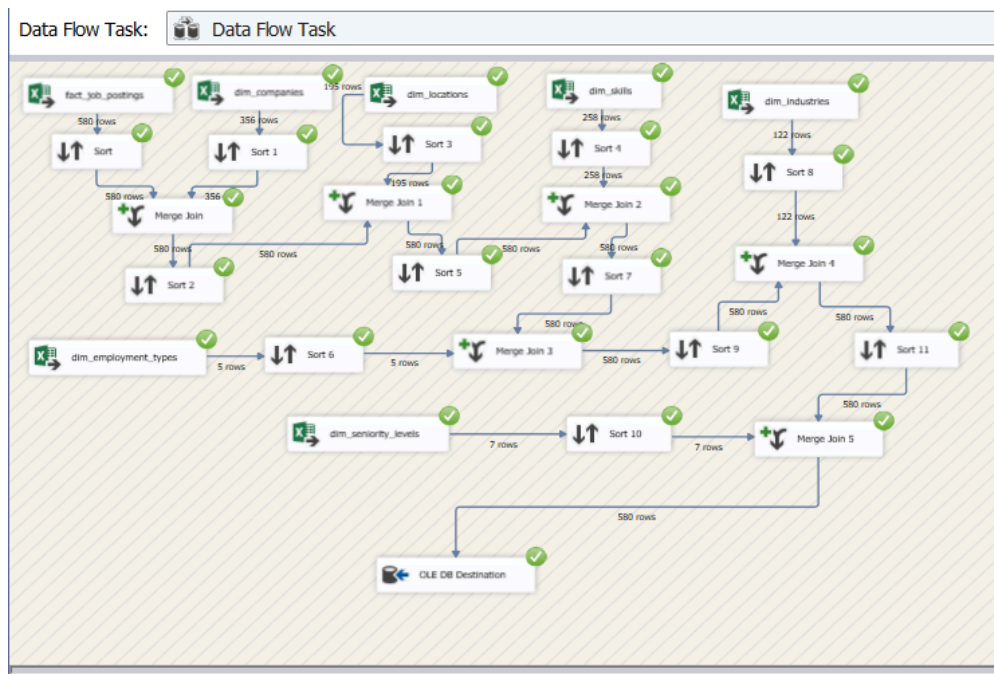
5.5.1. Tích hợp dữ liệu vào kho dữ liệu

5.5.1.1. Trích xuất dữ liệu (Extract)

- Dữ liệu được thu thập từ LinkedIn và các nguồn khác dưới dạng CSV.
- Sử dụng Python để làm sạch và chuẩn hóa dữ liệu.

5.5.1.2. Chuyển đổi dữ liệu (Transform)

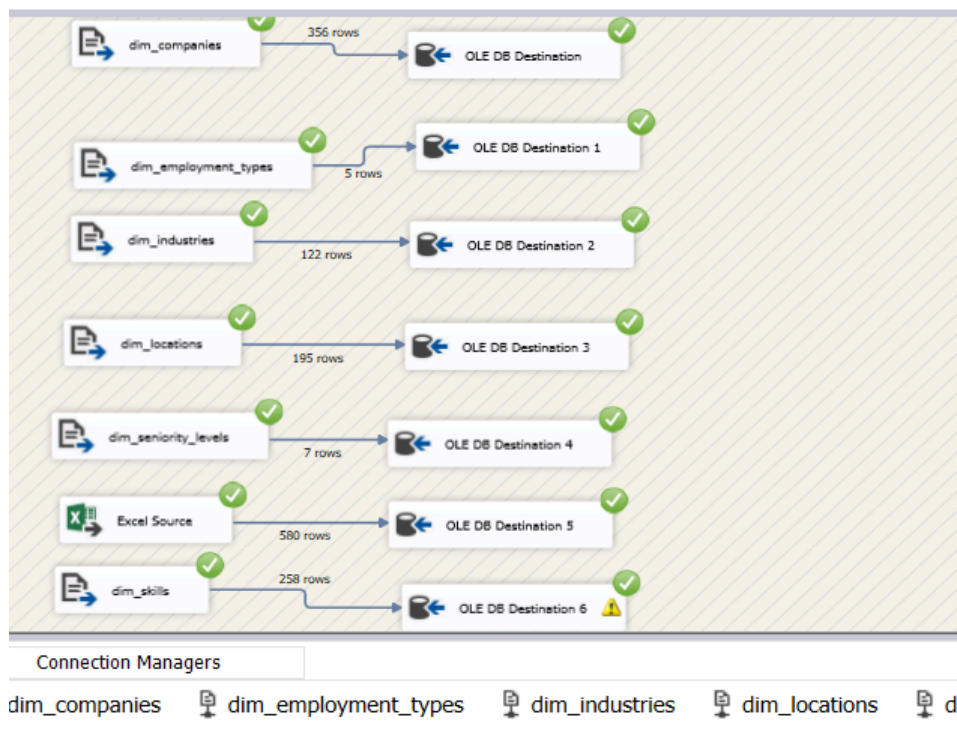
- Chuẩn hóa dữ liệu: Xóa trùng lặp, kiểu dữ liệu.
- Biến đổi các trường dữ liệu thành Dimension và Fact.



Hình 3: Luồng dữ liệu

5.5.1.3. Nạp dữ liệu (Load)

- Dữ liệu được nạp vào SQL Server.
- Sử dụng SSIS để tích hợp và đẩy dữ liệu vào kho.
- Kiểm tra tính đồng bộ và chất lượng dữ liệu sau khi load.



Hình 4: Kết quả nạp dữ liệu vào SQL

6. Khai thác kho dữ liệu

6.1. SQL Server Analysis Service (SSAS)

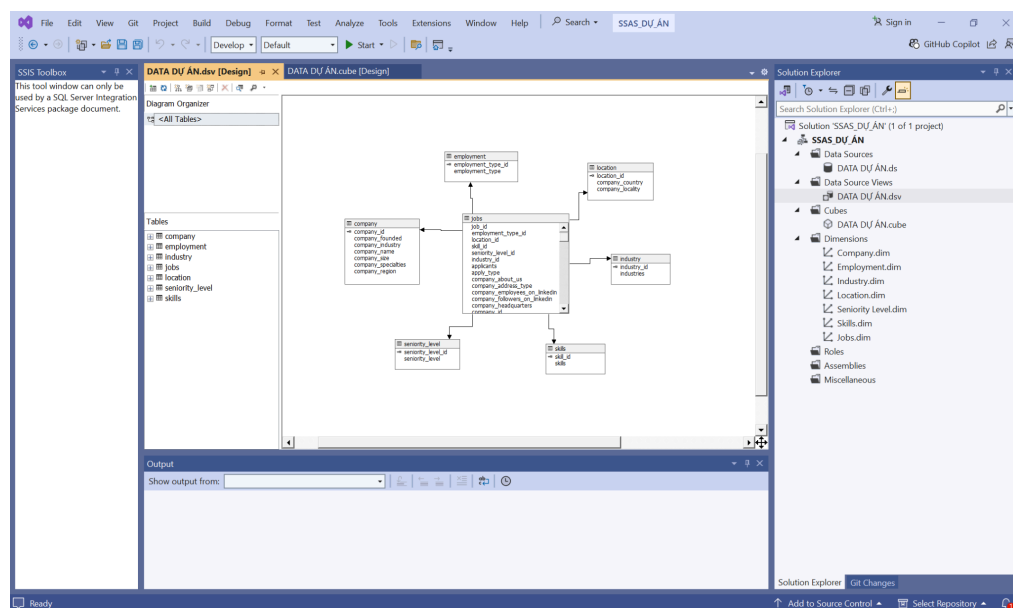
6.1.1 Giới thiệu về SSAS:

SQL Server Analysis Services (SSAS) là một công cụ mạnh mẽ trong hệ sinh thái Microsoft SQL Server, chuyên dùng để phân tích dữ liệu và hỗ trợ ra quyết định. SSAS giúp doanh nghiệp xây dựng các mô hình phân tích dữ liệu đa chiều (OLAP) hoặc mô hình bảng (Tabular) để tối ưu hóa việc truy vấn và phân tích dữ liệu lớn.

Các chức năng chính của SSAS bao gồm:

- Phân tích dữ liệu: SSAS cho phép xây dựng các mô hình phân tích dữ liệu giúp truy vấn và khai thác thông tin từ các nguồn dữ liệu lớn, hỗ trợ quá trình ra quyết định dựa trên dữ liệu thực tế.
- Xây dựng Cube OLAP: Với chế độ đa chiều (Multidimensional Mode), SSAS tổ chức dữ liệu theo dạng cube, tối ưu hóa khả năng truy vấn và phân tích các yếu tố phức tạp của dữ liệu.

6.2. SQL Server Analysis Service(SSAS) trong dự án



Hình 5: Mô hình dữ liệu trên SQL Server Analysis Services

6.3. Khai thác dữ liệu từ kho dữ liệu

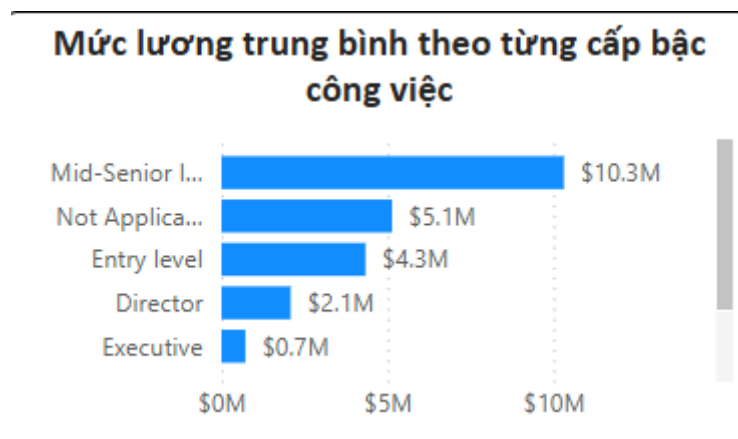
Dựa trên sơ đồ cơ sở dữ liệu đã có, ta có thể trích xuất thông tin theo các tiêu chí khác nhau như:

- Số lượng tin tuyển dụng theo ngành nghề

- Mức lương trung bình theo cấp bậc công việc
- Những kỹ năng phổ biến trong các tin tuyển dụng
- Số lượng ứng viên ứng tuyển theo từng công ty hoặc ngành
- Tình trạng tuyển dụng theo từng khu vực địa lý

6.4. Phân tích và tổng hợp dữ liệu

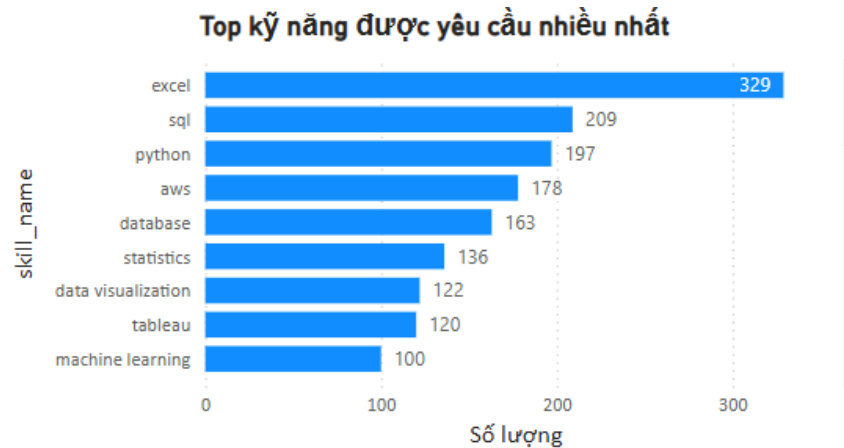
- Mức lương trung bình: cho từng cấp bậc, ngành nghề.



Hình 6: Mức lương trung bình theo các cấp bậc

- Tỷ lệ ứng viên trên mỗi tin tuyển dụng: để đánh giá mức độ hấp dẫn của công việc.

- Các kỹ năng: để xem kỹ năng nào được yêu cầu nhiều nhất.



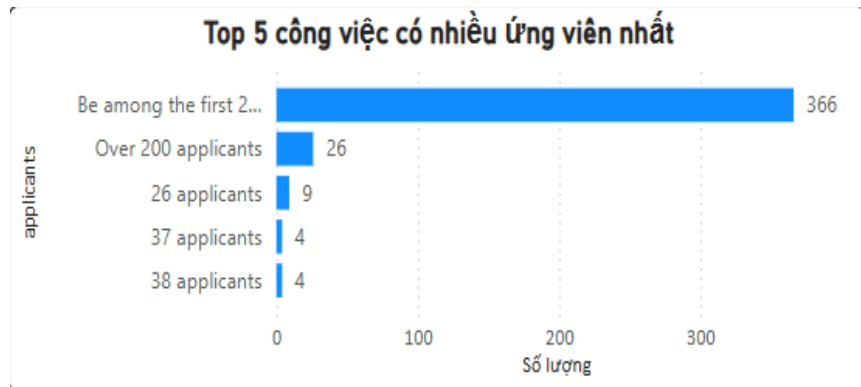
Hình 7: Các kỹ năng được yêu cầu nhiều nhất

- So sánh dữ liệu: So sánh mức lương giữa các ngành, khu vực.



Hình 8: Mức lương giữa các ngành

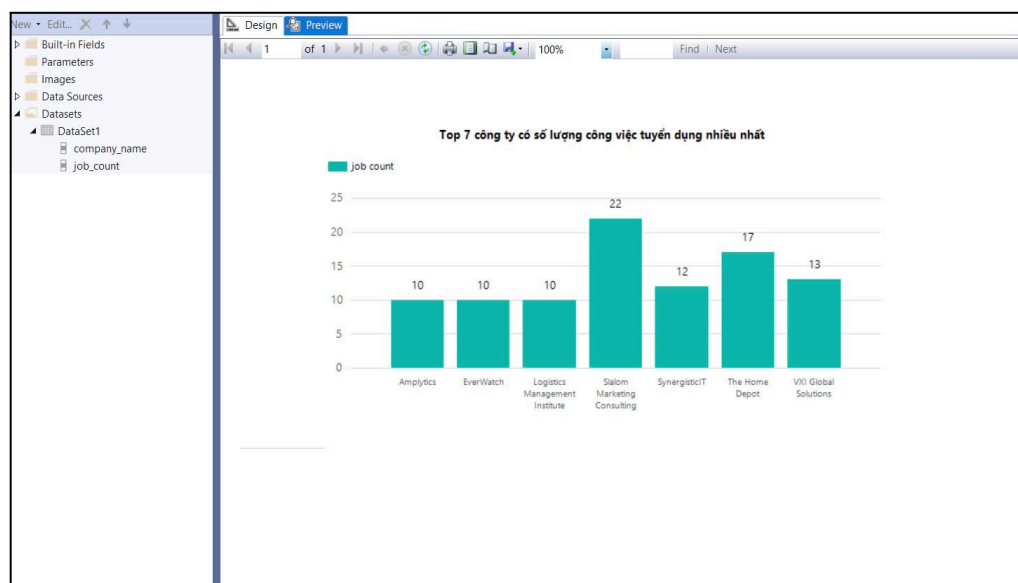
- Nhận diện xu hướng: Ngành nào đang có nhu cầu tuyển dụng cao nhất?
Cấp bậc nào đang có mức lương tăng trưởng?



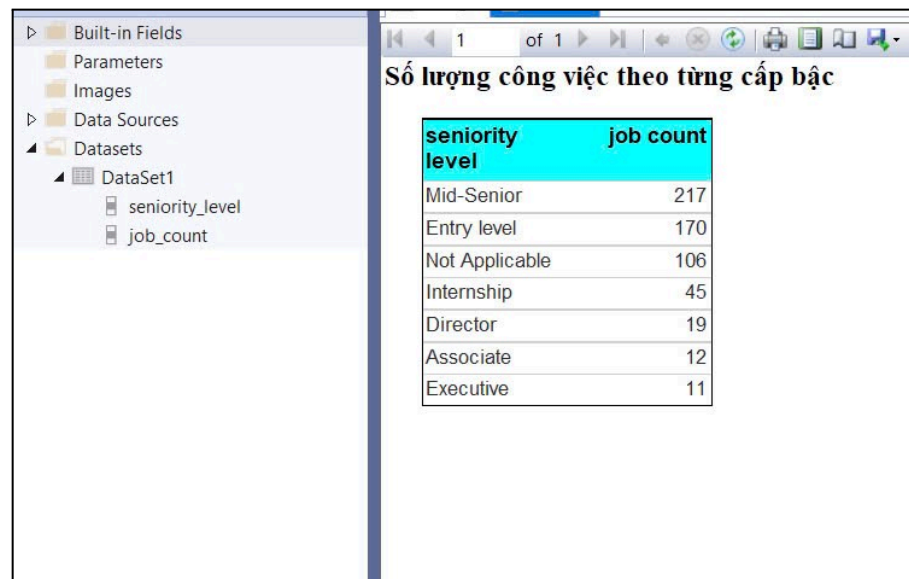
Hình 9: Các công việc có nhiều ứng viên nhất

6.5. Xây dựng báo cáo & trực quan hóa dữ liệu

6.5.1. SQL Server Reporting Services (SSRS) trong dự án



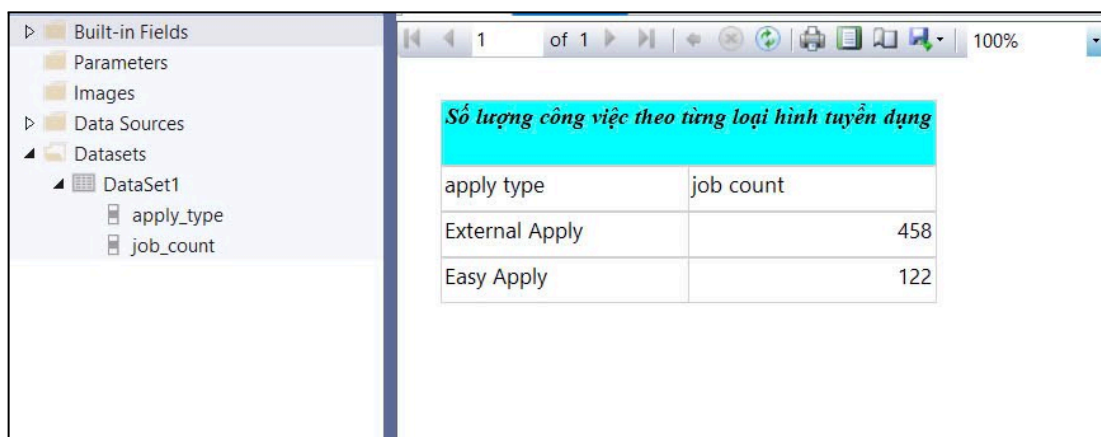
Hình 10: Công ty có số lượng công việc nhiều



The screenshot shows a BI tool interface with a left-hand pane containing a tree view of data sources. Under 'Data Sources', 'DataSets', and 'DataSet1', the fields 'seniority_level' and 'job_count' are listed. The main area displays a table titled 'Số lượng công việc theo từng cấp bậc' (Job count by seniority level). The table has two columns: 'seniority level' and 'job count'.

seniority level	job count
Mid-Senior	217
Entry level	170
Not Applicable	106
Internship	45
Director	19
Associate	12
Executive	11

Hình 11: Số lượng công việc theo các cấp bậc



The screenshot shows a BI tool interface similar to the previous one. The left-hand pane shows 'apply_type' and 'job_count' under 'DataSet1'. The main area displays a table titled 'Số lượng công việc theo từng loại hình tuyển dụng' (Job count by application type). The table has two columns: 'apply type' and 'job count'.

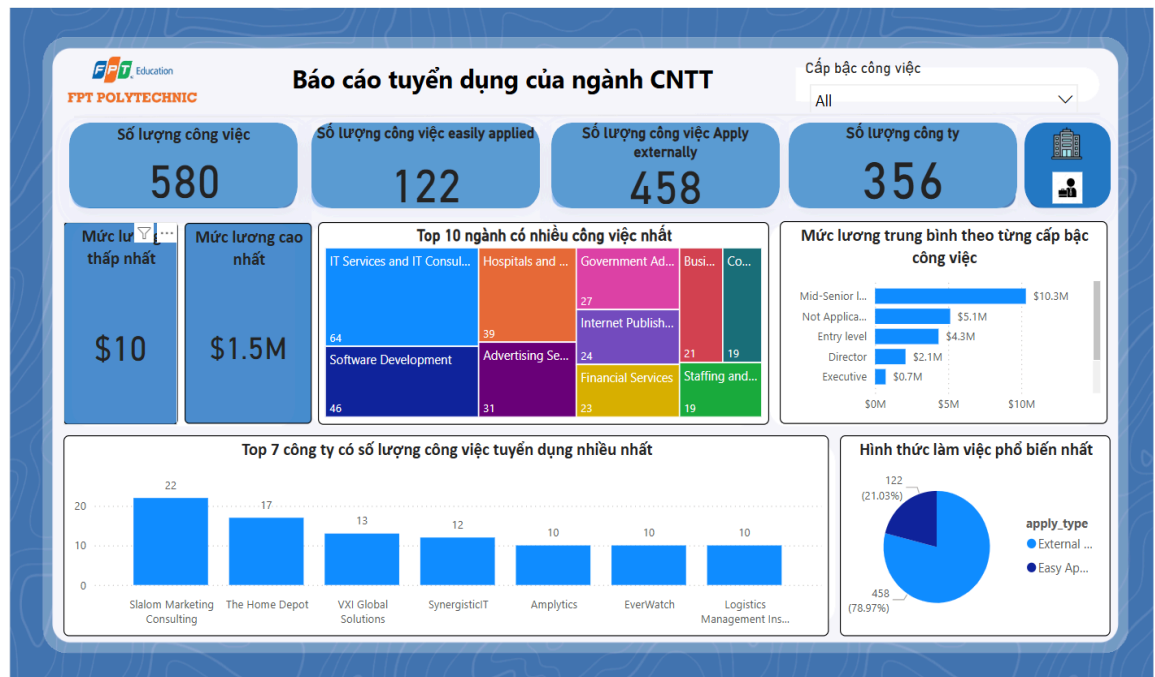
apply type	job count
External Apply	458
Easy Apply	122

Hình 12: Số lượng công việc theo hình thức tuyển dụng

6.5.2. Trực quan hóa dữ liệu (Power BI)

Sử dụng Power BI ta có thể trình bày báo cáo qua:

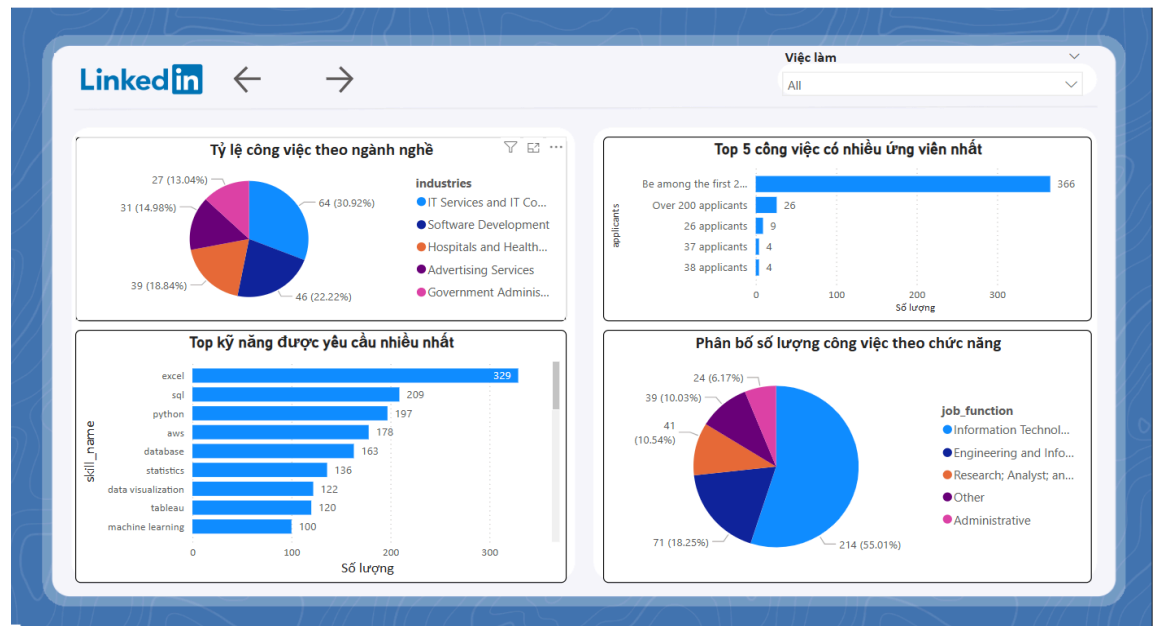
- Biểu đồ cột: Số lượng tin tuyển dụng theo ngành.
- Biểu đồ tròn: Phân bổ các kỹ năng yêu cầu.
- Biểu đồ Treemap: thể hiện số lượng công việc nhiều nhất



Hình 13 : Kết quả trực quan hóa trong Power BI (Page Overview)

Nhận xét:

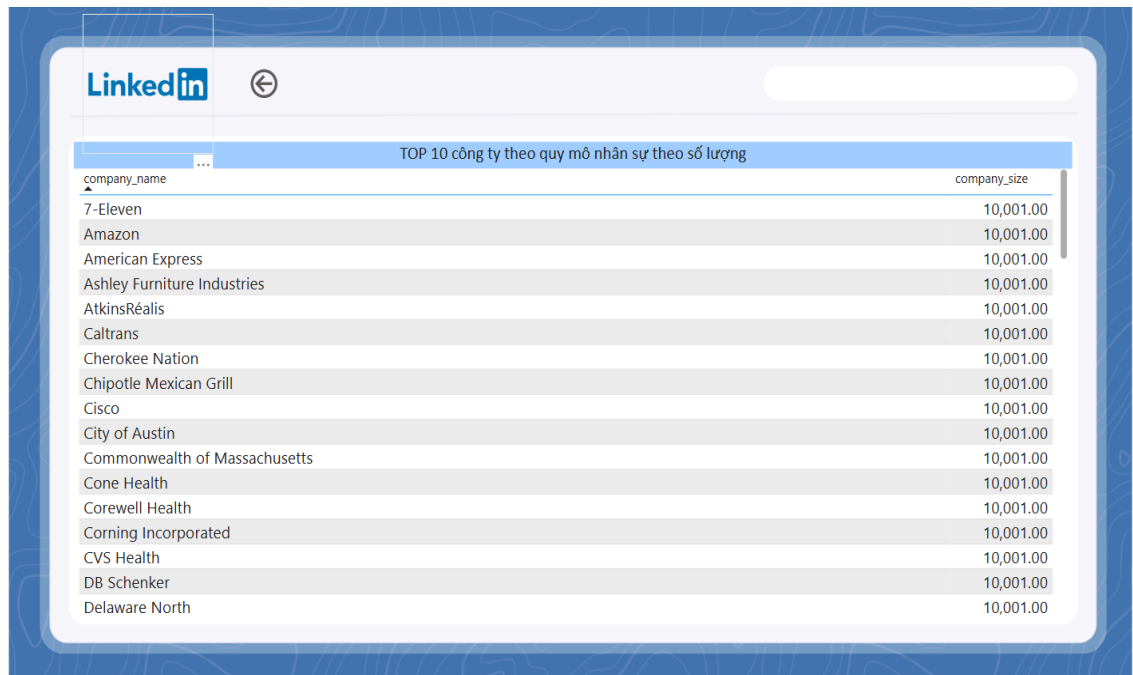
- Nhu cầu tuyển dụng ngành CNTT cao, đặc biệt trong Software Development, Marketing và Tài chính.
- Hình thức nộp đơn chủ yếu là Apply Externally, cho thấy nhiều công ty muốn ứng viên nộp đơn trực tiếp trên website của họ.
- Một số công ty tư vấn và công nghệ tuyển dụng nhiều nhất, phản ánh xu hướng doanh nghiệp đẩy mạnh số hóa.
- Mức lương trải rộng từ \$10 đến \$1.5M, có thể do sự khác biệt lớn giữa cấp bậc công việc hoặc dữ liệu chứa ngoại lệ.
- Cấp bậc công việc càng cao, mức lương càng lớn, đúng với xu hướng thị trường.



Hình 14: Page Nhân sự

Nhận xét:

- Tỷ lệ công việc theo ngành nghề: Ngành CNTT chiếm đa số, cho thấy nhu cầu nhân lực cao.
- Top công việc có nhiều ứng viên: Một số vị trí có hơn 300 ứng viên, cạnh tranh cao. Ứng viên cần chuẩn bị kỹ để nổi bật.
- Top kỹ năng được yêu cầu: SQL, Python, Excel đứng đầu, phản ánh nhu cầu lớn với phân tích dữ liệu và lập trình.
- Phân bố số lượng công việc: CNTT chiếm phần lớn, các vị trí như Kế toán, Hành chính ít hơn nhưng vẫn có nhu cầu.



The screenshot shows a LinkedIn interface with a table titled "TOP 10 công ty theo quy mô nhân sự theo số lượng". The table has two columns: "company_name" and "company_size". The data is as follows:

company_name	company_size
7-Eleven	10,001.00
Amazon	10,001.00
American Express	10,001.00
Ashley Furniture Industries	10,001.00
AtkinsRéalis	10,001.00
Caltrans	10,001.00
Cherokee Nation	10,001.00
Chipotle Mexican Grill	10,001.00
Cisco	10,001.00
City of Austin	10,001.00
Commonwealth of Massachusetts	10,001.00
Cone Health	10,001.00
Corewell Health	10,001.00
Corning Incorporated	10,001.00
CVS Health	10,001.00
DB Schenker	10,001.00
Delaware North	10,001.00

Hình 15: Page Công việc

Nhận xét:

- Báo cáo này cung cấp thông tin hữu ích về các công ty có quy mô nhân sự lớn nhất theo dữ liệu từ LinkedIn.

6.6. Đưa ra kết luận phân tích

Sau khi trực quan hóa, ta có thể rút ra:

- Ngành nào đang có nhu cầu tuyển dụng cao nhất?
- Mức lương trung bình của các ngành, cấp bậc?
- Những kỹ năng nào đang được yêu cầu nhiều nhất?
- Công ty nào có mức độ thu hút ứng viên cao nhất?
- Khu vực nào đang có nhiều cơ hội việc làm hơn?

7. Tổng kết

Dự án "Xây dựng kho dữ liệu tuyển dụng từ LinkedIn ngành CNTT" đã thành công trong việc thiết kế và triển khai một hệ thống kho dữ liệu giúp phân tích xu hướng tuyển dụng. Qua quá trình thu thập, xử lý và trực quan hóa dữ liệu, hệ thống đã hỗ trợ các doanh nghiệp, ứng viên và tổ chức giáo dục trong việc ra quyết định liên quan đến tuyển dụng và phát triển nghề nghiệp.

Các kết quả chính đạt được:

- Xây dựng một kho dữ liệu tập trung, lưu trữ thông tin tuyển dụng từ LinkedIn.
- Phân tích dữ liệu để nhận diện xu hướng tuyển dụng, kỹ năng phổ biến và mức lương trung bình.
- Trực quan hóa dữ liệu bằng Power BI, cung cấp các báo cáo trực quan và dễ hiểu.
- Hỗ trợ doanh nghiệp tối ưu hóa chiến lược tuyển dụng và giúp ứng viên định hướng nghề nghiệp.

Dự án này không chỉ giải quyết vấn đề phân tán dữ liệu tuyển dụng mà còn tạo ra một hệ thống có thể mở rộng, tích hợp với các nguồn dữ liệu khác trong tương lai. Với khả năng phân tích dữ liệu mạnh mẽ, kho dữ liệu này có tiềm năng hỗ trợ nhiều lĩnh vực khác ngoài tuyển dụng, như đào tạo nhân sự, dự báo thị trường lao động, và nghiên cứu phát triển ngành CNTT.

Qua quá trình thu thập, xử lý và trực quan hóa dữ liệu, hệ thống đã giúp giải quyết bài toán phân tán dữ liệu tuyển dụng trên LinkedIn, mang lại những lợi ích cụ thể như:

- Xây dựng kho dữ liệu tập trung, lưu trữ và quản lý thông tin tuyển dụng từ LinkedIn một cách hệ thống.
- Phân tích dữ liệu, nhận diện xu hướng tuyển dụng, kỹ năng phổ biến và mức lương trung bình trong ngành CNTT.
- Trực quan hóa dữ liệu bằng Power BI, cung cấp các báo cáo chi tiết, giúp người dùng dễ dàng theo dõi và đưa ra quyết định chính xác.
- Hỗ trợ doanh nghiệp tối ưu hóa chiến lược tuyển dụng, giúp xác định kỹ năng nào đang có nhu cầu cao, mức lương cạnh tranh và khu vực tuyển dụng tiềm năng.
- Giúp ứng viên định hướng nghề nghiệp, cung cấp thông tin về xu hướng việc làm, yêu cầu kỹ năng và mức lương trung bình để có kế hoạch học tập và phát triển phù hợp.
- Hỗ trợ các tổ chức giáo dục, cung cấp dữ liệu thực tế về nhu cầu nhân lực, giúp cập nhật chương trình đào tạo sát với yêu cầu của doanh nghiệp.

Hệ thống kho dữ liệu này không chỉ là một công cụ phân tích tuyển dụng mà còn có tiềm năng mở rộng để phục vụ các lĩnh vực khác như đào tạo nhân sự, dự báo thị trường lao động và nghiên cứu phát triển ngành CNTT.

Hạn chế của dự án

Mặc dù đạt được nhiều kết quả quan trọng, dự án vẫn còn một số hạn chế cần khắc phục để nâng cao hiệu quả và độ chính xác của hệ thống:

- Phạm vi dữ liệu còn hạn chế: Dữ liệu thu thập chủ yếu từ LinkedIn, chưa bao quát các nền tảng tuyển dụng khác như Indeed, VietnamWorks, TopCV, ITviec... Việc mở rộng phạm vi thu thập dữ liệu sẽ giúp cải thiện tính chính xác và độ bao phủ của hệ thống.

- Chưa áp dụng phương pháp học máy (Machine Learning): Hiện tại, hệ thống chủ yếu dựa vào phân tích dữ liệu thống kê. Việc tích hợp các mô hình Machine Learning có thể giúp dự đoán xu hướng tuyển dụng trong tương lai, từ đó cung cấp thông tin giá trị hơn cho doanh nghiệp và ứng viên.
- Chưa có hệ thống API mở: Việc phát triển API giúp hệ thống có thể tích hợp với các phần mềm quản lý nhân sự (HRM - Human Resource Management) của doanh nghiệp, giúp tự động hóa quy trình tuyển dụng và cập nhật dữ liệu liên tục.
- Chưa có chức năng cá nhân hóa: Hệ thống hiện tại chưa cung cấp các báo cáo tùy chỉnh theo nhu cầu của từng doanh nghiệp hoặc ứng viên. Việc phát triển các dashboard tùy chỉnh sẽ giúp người dùng khai thác dữ liệu hiệu quả hơn.

Thách thức về cập nhật dữ liệu theo thời gian thực: Việc thu thập và cập nhật dữ liệu hiện tại vẫn chưa hoàn toàn theo thời gian thực. Việc cải thiện hệ thống để tự động hóa quá trình cập nhật sẽ giúp đảm bảo dữ liệu luôn mới nhất và chính xác nhất.

Hướng phát triển trong tương lai

Đề nâng cao hiệu suất và mở rộng ứng dụng của hệ thống, dự án có thể được phát triển theo các hướng sau:

- Mở rộng phạm vi thu thập dữ liệu: Tích hợp thêm các nguồn dữ liệu tuyển dụng từ các nền tảng khác như Indeed, VietnamWorks, ITviec, TopDev... để tăng độ chính xác và tính bao quát của kho dữ liệu.
- Phát triển tính năng cá nhân hóa:

- + Cung cấp báo cáo tùy chỉnh theo từng doanh nghiệp, giúp họ có cái nhìn chi tiết hơn về thị trường lao động trong lĩnh vực của mình.
- + Tạo bảng điều khiển (dashboard) cá nhân cho ứng viên, giúp họ theo dõi xu hướng việc làm và đề xuất lộ trình phát triển kỹ năng.
- + Tối ưu hóa hệ thống cập nhật dữ liệu theo thời gian thực: Cải thiện hạ tầng công nghệ để đảm bảo dữ liệu luôn được cập nhật nhanh chóng và chính xác nhất.
- Phát triển mô-đun đào tạo và tư vấn nghề nghiệp: Kết hợp dữ liệu phân tích với các khóa đào tạo trực tuyến, giúp ứng viên nâng cao kỹ năng cần thiết để đáp ứng nhu cầu thị trường.

Tổng kết

Dự án "Xây dựng kho dữ liệu tuyển dụng từ LinkedIn ngành CNTT" đã đặt nền móng quan trọng trong việc ứng dụng dữ liệu lớn vào quản lý nhân sự và tuyển dụng. Không chỉ giúp doanh nghiệp tối ưu hóa chiến lược tuyển dụng, hệ thống còn hỗ trợ ứng viên và tổ chức giáo dục trong việc định hướng phát triển nghề nghiệp.

Với tiềm năng mở rộng và cải tiến trong tương lai, hệ thống này có thể trở thành một công cụ mạnh mẽ không chỉ trong lĩnh vực tuyển dụng mà còn trong quản lý nhân sự, đào tạo và dự báo thị trường lao động. Mở rộng dữ liệu và cải thiện khả năng cá nhân hóa sẽ giúp nâng cao giá trị của dự án, đóng góp tích cực vào sự phát triển của ngành CNTT nói riêng và thị trường lao động nói chung.

8. Tài liệu tham khảo



Data



Trello



Canva