

Make My Career

Description

Searching for a new job is a huge task. Making a career change is monumental. When you want a fresh start or a change in your career, you will need help in the form of advice and an outside perspective. When you want to go beyond your current role, and explore new roles, you can achieve it by simply adding a few more skills to the ones you already have under your belt. We have a system that will help you achieve this change.

MakeMyCareer is a system that will provide suggestions to a user in the form of prospective careers, positions, skills, and resources to add to their profile based on their current skills, expectations, and the job market. Also, the user will be able to make their profile impressive and enticing to prospective recruiters in terms of their resume, and their online presence for an easy hiring process.

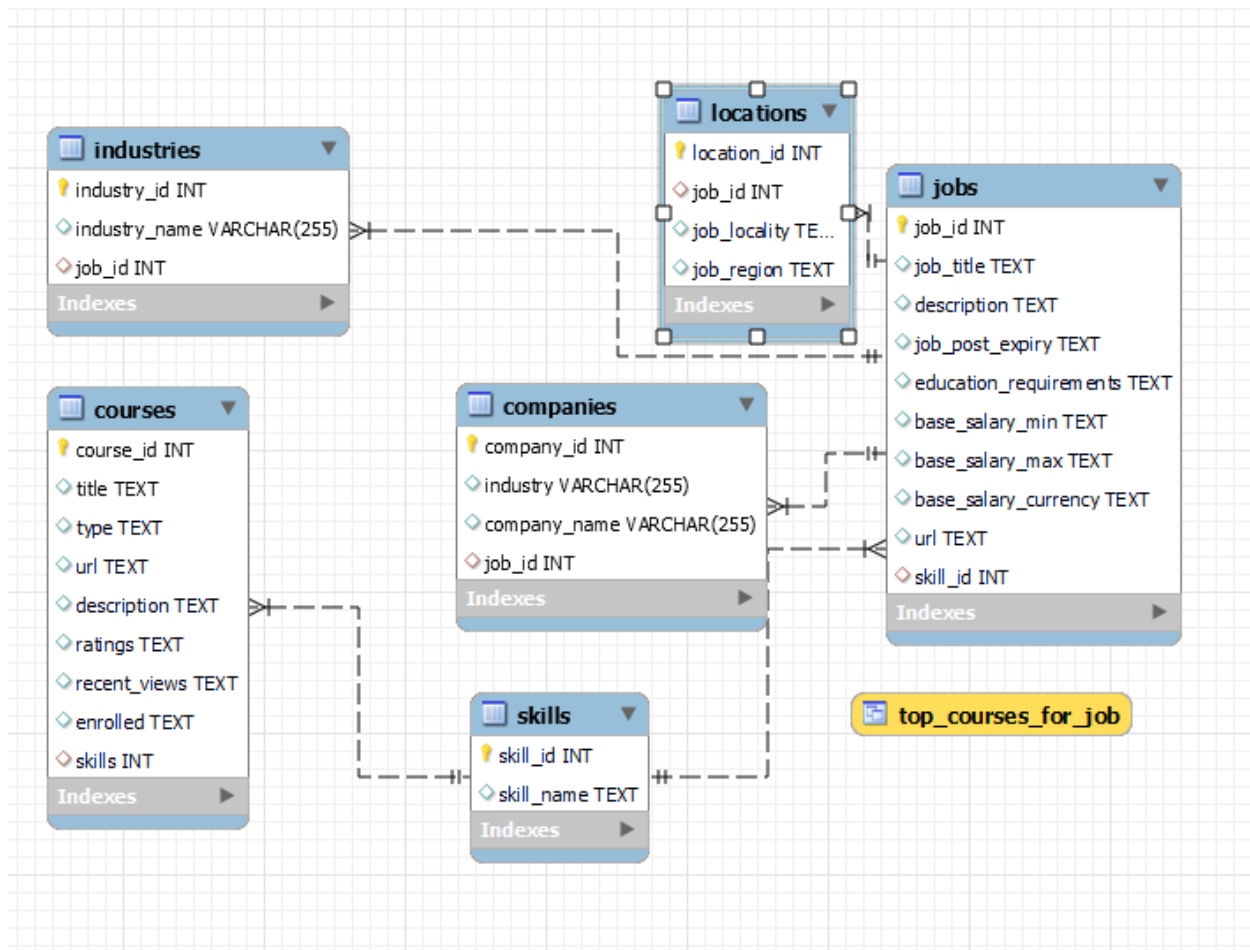
The System will provide -

- Prospective Positions or Profiles based on the current skills a user already has, their work experience, expectations, and the demand in the job market. These expectations can include requirements like salary, work-life balance, and ideologies.
- Based on a Target Position, the user will want to know any additional skills missing from their profile. These will be looked up and filtered using the latter parameters, and the user's target companies.
- Once the user knows the skills required, they would need trending and quality resources to learn these skills.

Team:

Shabina Singh singh.shab@northeastern.edu 002652525

Entity Relationship Diagram



Sample Data

Jobs :

| job_id | job_title | description | job_post | education_r | base_si | base_ | base_ | url | skill_id |
|--------|------------------------------|--|----------|--------------|---------|-------|-------|--|----------|
| 0 | Data Scientist Internship | iheartmediacurrent employee contingent ... | 2022-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 0 |
| 1 | Data Scientist Internship | iheartmediacurrent employee contingent ... | 2022-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 16 |
| 2 | Data Scientist Internship | iheartmediacurrent employee contingent ... | 2022-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 18 |
| 3 | Data Scientist Intern | task internship title interndept name risk ... | 2022-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 0 |
| 4 | Data Scientist Internship... | 2022 75202asurion science internship or... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 0 |
| 5 | Data Scientist Internship... | 2022 75202asurion science internship or... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 4 |
| 6 | Data Scientist Internship... | 2022 75202asurion science internship or... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 7 |
| 7 | Data Scientist Internship... | 2022 75202asurion science internship or... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 15 |
| 8 | Data Scientist Internship... | 2022 75202asurion science internship or... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 16 |
| 9 | Data Scientist Internship... | 2022 75205asurion science internship ml... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 0 |
| 10 | Data Scientist Internship... | 2022 75205asurion science internship ml... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 4 |
| 11 | Data Scientist Internship... | 2022 75205asurion science internship ml... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 7 |
| 12 | Data Scientist Internship... | 2022 75205asurion science internship ml... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 15 |
| 13 | Data Scientist Internship... | 2022 75205asurion science internship ml... | 2023-... | bachelor ... | NULL | NULL | NULL | https://www.linkedin.com/jobs/view/data-sci... | 16 |

Courses :

| course_id | title | type | url | description | ratings | recent_views | enrolled | skills |
|-----------|--------------------------|--------|---|---------------------|-------------|---------------------|-------------------------|--------|
| 0 | Accounting for Merg... | Course | https://www.coursera.org/learn/accounting-fo... | course aim assi... | NULL | 9,141 recent views | NULL | 4 |
| 1 | Assisting Public Sect... | Course | https://www.coursera.org/learn/assist-public-s... | develop data a... | NULL | 7,607 recent views | NULL | 4 |
| 2 | Assisting Public Sect... | Course | https://www.coursera.org/learn/assist-public-s... | develop data a... | NULL | 7,607 recent views | NULL | 5 |
| 3 | Atendã'n prehospit... | Course | https://www.coursera.org/learn/ictus-agudo-e... | el ictus e una e... | 231 ratings | 3,625 recent views | 4,251 already enrolled | 5 |
| 4 | Atendã'n prehospit... | Course | https://www.coursera.org/learn/ictus-agudo-e... | el ictus e una e... | 231 ratings | 3,625 recent views | 4,251 already enrolled | 17 |
| 5 | Build, Train, and De... | Course | https://www.coursera.org/learn/ml-pipelines-bert | second course ... | 115 ratings | 17,709 recent views | 9,797 already enrolled | 4 |
| 6 | Build, Train, and De... | Course | https://www.coursera.org/learn/ml-pipelines-bert | second course ... | 115 ratings | 17,709 recent views | 9,797 already enrolled | 5 |
| 7 | Build, Train, and De... | Course | https://www.coursera.org/learn/ml-pipelines-bert | second course ... | 115 ratings | 17,709 recent views | 9,797 already enrolled | 15 |
| 8 | Build, Train, and De... | Course | https://www.coursera.org/learn/ml-pipelines-bert | second course ... | 115 ratings | 17,709 recent views | 9,797 already enrolled | 17 |
| 9 | Build, Train, and De... | Course | https://www.coursera.org/learn/ml-pipelines-bert | second course ... | 115 ratings | 17,709 recent views | 9,797 already enrolled | 18 |
| 10 | Comercio, Inmigrad... | Course | https://www.coursera.org/learn/comercio-inmi... | este e el segu... | 77 ratings | 3,112 recent views | 2,901 already enrolled | 5 |
| 11 | FPGA computing sys... | Course | https://www.coursera.org/learn/fpga-intro | course anyone... | 224 ratings | 7,362 recent views | 13,249 already enrolled | 4 |

Skills :

| skill_id | skill_name |
|----------|------------|
| 0 | django |
| 1 | php |
| 2 | java |
| 3 | numpy |
| 4 | ai |
| 5 | ui |
| 6 | mysql |
| 7 | flask |
| 8 | pytorch |
| 9 | spark |
| 10 | tensorflow |
| 11 | tableau |

Locations :

| | location_id | job_id | job_locality | job_region |
|---|-------------|--------|------------------|---------------|
| ▶ | 0 | 0 | Texas | United States |
| | 1 | 1 | Farmington Hills | MI |
| | 2 | 2 | Nashville | TN |
| | 3 | 3 | Nashville | TN |
| | 4 | 4 | Home | MN |
| | 5 | 5 | Wayne | NJ |
| | 6 | 6 | Parsippany | NJ |
| | 7 | 7 | Raleigh | NC |
| | 8 | 8 | Phoenix | AZ |
| | 9 | 9 | Los Angeles | CA |
| | 10 | 10 | Home | PA |
| | 11 | 11 | Los Angeles | CA |

Companies :

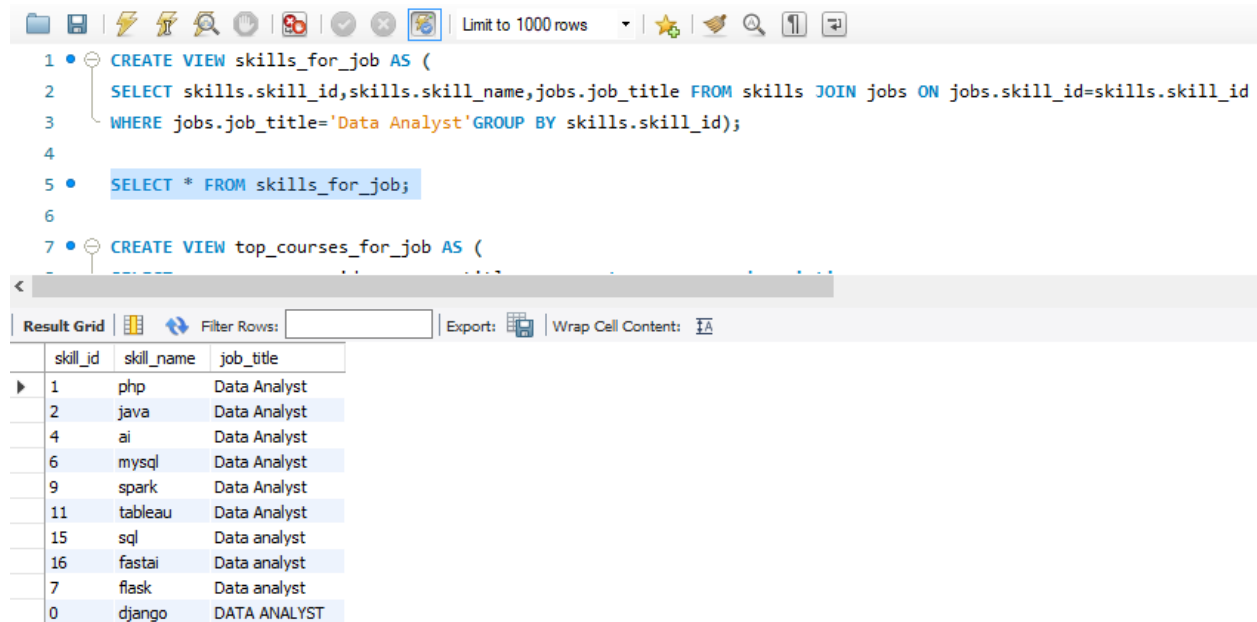
| | company_name |
|---|---|
| ▶ | iHeartMedia |
| | Mercedes-Benz Financial Services USA LLC |
| | Asurion |
| | Digi-Key Electronics |
| | Valley Bank |
| | Langan Engineering & Environmental Services |
| | IBM |
| | Amgen |
| | Highmark Health |
| | BioSpace |
| | Dana-Farber Cancer Institute |
| | USAA |

Industries

| | industry_name |
|---|--|
| ▶ | Radio and Television Broadcasting |
| | Financial Services |
| | Insurance |
| | Appliances and Electrical and Electronics Manuf... |
| | Banking |
| | Civil Engineering |
| | Computer Hardware Manufacturing |
| | Biotechnology Research |
| | Pharmaceutical Manufacturing |
| | Hospitals and Health Care |
| | Online Audio and Video Media |
| | Technology |

Views based on Use Cases

Find skills required for a particular Job Position -



```
1 CREATE VIEW skills_for_job AS (  
2   SELECT skills.skill_id,skills.skill_name,jobs.job_title FROM skills JOIN jobs ON jobs.skill_id=skills.skill_id  
3   WHERE jobs.job_title='Data Analyst'GROUP BY skills.skill_id);  
4  
5 SELECT * FROM skills_for_job;  
6  
7 CREATE VIEW top_courses_for_job AS (  
8   SELECT courses.course_id, courses.title,courses.type,courses.description,  
9   courses.ratings,courses.recent_views,courses.enrolled,skills.skill_name,jobs.job_title  
10  FROM courses JOIN (skills JOIN jobs ON jobs.skill_id=skills.skill_id ) ON courses.skills=skills.skill_id  
11  WHERE jobs.job_title in ('Data Analyst') ORDER BY courses.ratings);  
12 SELECT * FROM top_courses_for_job;
```

Result Grid

| | skill_id | skill_name | job_title |
|---|----------|------------|--------------|
| ▶ | 1 | php | Data Analyst |
| | 2 | java | Data Analyst |
| | 4 | ai | Data Analyst |
| | 6 | mysql | Data Analyst |
| | 9 | spark | Data Analyst |
| | 11 | tableau | Data Analyst |
| | 15 | sql | Data analyst |
| | 16 | fastai | Data analyst |
| | 7 | flask | Data analyst |
| | 0 | django | DATA ANALYST |

Find top courses for my job position -

```
6 CREATE VIEW top_courses_for_job AS (  
7   SELECT courses.course_id, courses.title,courses.type,courses.description,  
8   courses.ratings,courses.recent_views,courses.enrolled,skills.skill_name,jobs.job_title  
9   FROM courses JOIN (skills JOIN jobs ON jobs.skill_id=skills.skill_id ) ON courses.skills=skills.skill_id  
10  WHERE jobs.job_title in ('Data Analyst') ORDER BY courses.ratings);  
11  
12 SELECT * FROM top_courses_for_job;
```

Result Grid

| | course_id | title | type | description | ratings | recent_views | enrolled | skill_name | job_title |
|---|-----------|--------------------------|--------|-----------------------------------|---------|---------------------|------------------------|------------|--------------|
| ▶ | 1316 | Architecting Solution... | Course | looking get technical looking ... | NULL | 51,491 recent views | 2,230 already enrolled | ai | Data Analyst |
| | 320 | Advanced Data Mod... | Course | develop working knowledge f... | NULL | 4,923 recent views | NULL | mysql | Data Analyst |
| | 1295 | Aprendizado de mÃi... | Course | este curso mergulha no fund... | NULL | 2,221 recent views | NULL | ai | Data Analyst |
| | 1341 | Arquitecturas de Big ... | Course | el curso de arquiteturas de ... | NULL | 2,002 recent views | NULL | spark | Data Analyst |
| | 924 | Advanced Data Mod... | Course | develop working knowledge f... | NULL | 4,923 recent views | NULL | mysql | Data Analyst |
| | 1310 | Architecting Google ... | Course | questo curso architecting go... | NULL | NULL | NULL | ai | Data Analyst |
| | 1311 | Architecting Google ... | Course | questo curso architecting go... | NULL | NULL | NULL | ai | Data Analyst |
| | 1313 | Architecting Google ... | Course | questo curso architecting go... | NULL | NULL | NULL | ai | Data Analyst |
| | 1437 | Attracting and Sourc... | Course | shortlisting interviewing proc... | NULL | 1,750 recent views | NULL | ai | Data Analyst |
| | 1388 | Arilha Networkinn Ba... | Course | arilha networkinn essential v... | NULL | 1.889 recent views | NULL | ai | Data Analyst |

Find jobs for set of skills in user's profile-

```

13 • CREATE VIEW jobs_for_skill AS (
14   SELECT courses.course_id, courses.title, courses.type, courses.description,
15   courses.ratings, courses.recent_views, courses.enrolled, skills.skill_name, jobs.job_title
16   FROM courses JOIN (skills JOIN jobs ON jobs.skill_id=skills.skill_id )
17   ON courses.skills=skills.skill_id WHERE skills.skill_name in ('ai', 'python', 'sql')
18   GROUP BY skills.skill_name);
19
20 • SELECT * FROM top_courses_for_job;

```

| course_id | title | type | description | ratings | recent_views | enrolled | skill_name | job_title |
|-----------|--------------------------|--------|-----------------------------------|---------|---------------------|------------------------|------------|--------------|
| 1316 | Architecting Solution... | Course | looking get technical looking ... | NULL | 51,491 recent views | 2,230 already enrolled | ai | Data Analyst |
| 320 | Advanced Data Mod... | Course | develop working knowledge f... | NULL | 4,923 recent views | NULL | mysql | Data Analyst |
| 1295 | Aprendizado de mÃi... | Course | este curso mergulha no fund... | NULL | 2,221 recent views | NULL | ai | Data Analyst |
| 1341 | Arquitecturas de Big ... | Course | el curso de arquiteturas de ... | NULL | 2,002 recent views | NULL | spark | Data Analyst |
| 924 | Advanced Data Mod... | Course | develop working knowledge f... | NULL | 4,923 recent views | NULL | mysql | Data Analyst |
| 1310 | Architecting Google ... | Course | questo curso architecting go... | NULL | NULL | NULL | ai | Data Analyst |
| 1311 | Architecting Google ... | Course | questo curso architecting go... | NULL | NULL | NULL | ai | Data Analyst |
| 1313 | Architecting Google ... | Course | questo curso architecting go... | NULL | NULL | NULL | ai | Data Analyst |
| 1437 | Attracting and Sourc... | Course | shortlisting interviewing proc... | NULL | 1,750 recent views | NULL | ai | Data Analyst |
| 1388 | Aruba Networking Ba... | Course | aruba networking essential v... | NULL | 1,889 recent views | NULL | ai | Data Analyst |

Which courses are available on AI which provide a certificate?

```

22
23 • CREATE VIEW cert_courses_for_skill AS (
24   SELECT * FROM courses JOIN skills ON courses.skills=skills.skill_id
25   WHERE skills.skill_name='ai' and type = 'Course');
26
27 • SELECT * FROM cert_courses_for_skill;
28

```

| course_id | title | type | url | description | ratings | recent_views | enrolled | skills | skill_id | skill_nar |
|-----------|---|--------|-------------------------|----------------------------|-------------|---------------------|---------------|--------|----------|-----------|
| 0 | Accounting for Mergers and Acquisitions: Foun... | Course | https://www.coursera... | course aim assisting in... | NULL | 9,141 recent views | NULL | 4 | 4 | ai |
| 1 | Assisting Public Sector Decision Makers With Pol... | Course | https://www.coursera... | develop data analysis ... | NULL | 7,607 recent views | NULL | 4 | 4 | ai |
| 5 | Build, Train, and Deploy ML Pipelines using BER... | Course | https://www.coursera... | second course practic... | 115 ratings | 17,709 recent views | 9,797 alre... | 4 | 4 | ai |
| 11 | FPGA computing systems: Background knowled... | Course | https://www.coursera... | course anyone passio... | 224 ratings | 7,362 recent views | 13,249 alr... | 4 | 4 | ai |
| 13 | Future Healthcare Payment Models Coursera | Course | https://www.coursera... | course review driver h... | 34 ratings | 3,911 recent views | 3,719 alre... | 4 | 4 | ai |
| 14 | Improving Immunity Based on Traditional Easte... | Course | https://www.coursera... | exercise medicine regu... | 61 ratings | 6,906 recent views | 3,718 alre... | 4 | 4 | ai |
| 17 | Incrementar - Parte 2 y Controlar Coursera | Course | https://www.coursera... | el curso explora herra... | NULL | 5,330 recent views | NULL | 4 | 4 | ai |
| 19 | IntelÂ® Network Academy - Network Transfor... | Course | https://www.coursera... | welcome intelâ network... | 473 ratings | 20,587 recent views | 29,575 alr... | 4 | 4 | ai |
| 20 | Introduction to Academic Writing Coursera | Course | https://www.coursera... | welcome introduction ... | NULL | 19,604 recent views | NULL | 4 | 4 | ai |
| 21 | Know Thyself - The Value and Limits of Self-Kno... | Course | https://www.coursera... | according legend inscri... | 507 ratings | 32,696 recent views | 51,281 alr... | 4 | 4 | ai |
| 22 | Les bits et les octets des rÃ@seaux informatiq... | Course | https://www.coursera... | ce cours est conã u po... | NULL | 4,896 recent views | NULL | 4 | 4 | ai |
| 24 | Mandarin Chinese for Intermediate Learners: P... | Course | https://www.coursera... | mandarin chinese 1 chi... | 102 ratings | 4,579 recent views | 9,988 alre... | 4 | 4 | ai |

Which is the most trending course on mysql?

```

29 • CREATE VIEW courses_for_skill AS (
30   SELECT * FROM courses JOIN skills ON courses.skills=skills.skill_id
31   WHERE skills.skill_name='ai' ORDER BY recent_views DESC);
32
33 • SELECT * FROM courses_for_skill LIMIT 1;

```

| Result Grid | | | | | | | | | | |
|--|-------------------------|--------|----------------------|------------------|---------------|---------------------|--------------------------|--------|----------|------------|
| Filter Rows: <input type="text"/> | | | | | | | | | | |
| Export: Wrap Cell Content: Fetch rows: | | | | | | | | | | |
| course_id | title | type | url | description | ratings | recent_views | enrolled | skills | skill_id | skill_name |
| ▶ 1003 | Advanced Styling wit... | Course | https://www.cours... | used case eve... | 4,297 ratings | 99,922 recent views | 102,646 already enrolled | 4 | 4 | ai |

Scripts used :

https://github.com/anibahs/MakeMyCareer/blob/main/FinalProject/mmc_views.sql

Use Cases

1. Find skills required for a Data Analyst Job Position
2. Find courses for my job position
3. Find jobs for set of skills in user's profile
4. How many paid courses are available on AI which provide a certificate?
5. Which is the most trending course on Natural Language Processing?

Queries -

```
query1 = "SELECT * \
FROM skills JOIN jobs ON jobs.skills=skills.skill_id \
WHERE jobs.job_title='Data Analyst' GROUP BY skills.skill_id;"
query2 = "SELECT * \
FROM courses JOIN (skills JOIN jobs ON jobs.skills=skills.skill_id) \
ON courses.skills=skills.skill_id WHERE jobs.job_title in ('Data Analyst') \
GROUP BY courses.skills ORDER BY courses.ratings"
query3 = "SELECT * \
FROM jobs JOIN skills ON jobs.skills=skills.skill_id \
WHERE skills.skill_name in ('ai','python', 'sql') GROUP BY skills.skill_name; "
query4 = "SELECT * \
FROM courses JOIN skills ON courses.skills=skills.skill_id \
WHERE skills.skill_name='ai' and type = 'Professional Certificate';"
query5 = "SELECT * FROM courses JOIN skills ON courses.skills=skills.skill_id \
WHERE skills.skill_name='ai' ORDER BY recent_views DESC LIMIT 1;"
```

Twitter bot use cases

Use Case: User can look for opening for their target job position Description: User can look for opening for a position named "Engineer"

Actor: User

Precondition: User should have a valid target position name

Steps:

Actor action: User request for list of job openings for his target position.

System Responses: If the position exists, the system will return a list of job openings posted.

Post Condition: List of job openings suggested

Alternate Path: The user request is not correct and system throws an error

Error: User information is incorrect

Use Case: User can look for openings posted by their dream company handle
Description: Search for job posts posted by a particular user

Actor: User

Precondition: User should have a company name user is target

Steps:

Actor action: User request for list of job openings for his target position.

System Responses: If the company has posted job openings, the system will return the list.

Post Condition: List of job openings suggested

Alternate Path: The user request is not correct and system throws an error

Error: User information is incorrect

Use Case: User can look for openings posted within last 5 days and for a particular position
Description: Search for job posts posted within last 5 days

Actor: User

Precondition: User should have a valid target position name

Steps:

Actor action: User request for list of job openings for his target position.

System Responses: The system will return a list of job posts.

Post Condition: List of job openings suggested

Alternate Path: The user request is not correct and system throws an error

Error: User information is incorrect

Use Case: User can assess which job positions are more in demand
Description: Search for job posts for different job positions

Actor: User

Precondition: User should have a valid target position name

Steps:

Actor action: User request for list of job openings for his target position.

System Responses: The system will return a count of job posts for a position.

Post Condition: List of job openings suggested

Alternate Path: The user request is not correct and system throws an error

Error: User information is incorrect

Use Case: User can assess which companies are posting more jobs

Description: Search for job posts for job positions by different companies

Actor: User

Precondition: User should have a valid target position name and target company

Steps:

Actor action: User request for list of job openings for his target position posted by company.

System Responses: The system will return a count of job posts posted by company handle.

Post Condition: Count of job openings suggested

Alternate Path: The user request is not correct and system throws an error

Error: User information is incorrect

File: <https://github.com/anibahs/MakeMyCareer/blob/main/FinalProject/Scripts/TwitterBot.ipynb>

Steps

Sources :

For this project, one of the top websites which could provide a substantial amount and clean information on job openings and additional information related to job searches was LinkedIn and was scraped to obtain a snapshot of data on current job openings. Also Coursera was used to scrape data on skills and resources.

Scraping:

Two python scripts were used along with BeautifulSoup library to download data from LinkedIn and Coursera directories, hit the subsequent page for each job/course and extract information from each HTML based on the right tags. This data was stored in their respective CSV files to be cleaned by a different python script.

Files:

<https://github.com/anibahs/MakeMyCareer/blob/main/FinalProject/Scripts/CourseraScraper.ipynb>
<https://github.com/anibahs/MakeMyCareer/blob/main/FinalProject/Scripts/LinkedInScraper.ipynb>

Data Cleaning:

Another python script along with pandas, numpy and matplotlib and other libraries was used to clean the data.

It ensured the data does not contain null values - empty values were converted to Nan and completed empty rows were deleted.

Any data in different languages was removed.

Multi-valued entries were removed by creating multiple entries for each attribute. In jobs data, industries, companies and skills were found to be multivalued. In courses, skills were multivalued.

Data Quality Analysis:

Data visualization in the form of countplot were created to check if the number of jobs were accurate and consistent with the industry and skills in the job market. Data completeness was ensured by removing null values and multivalues.

File -

<https://github.com/anibahs/MakeMyCareer/blob/main/FinalProject/Scripts/DataExtraction.ipynb>

Data Extraction:

Using a sample set of technical skills the model MultinomialNB was trained on this data to predict additional technical skills and some of the top skills were used to relate to the job positions.

Database Creation:

All pandas dataframes on jobs, courses, skills, and locations were formed, cleaned and also linked with relevant attributes.

Tables were created using SQL queries with all the necessary constraints for primary and foreign key relationships.

This data was then inserted into a database on MySQL Workbench through a connection with python using mysql libraries..

Normalization:

Once the database was set up, the data was found to be already in First Normal Form. It was then converted into Second Normal Form by moving data with partial dependencies (company, industry) into new tables . Location and Skills were already removed and linked with foreign keys.

It was then converted into 3rd normal form by removing redundant data from the tables and eliminating transitive dependencies.

SQL scripts used -

<https://github.com/anibahs/MakeMyCareer/blob/main/FinalProject/Scripts/1NF.sql>

<https://github.com/anibahs/MakeMyCareer/blob/main/FinalProject/Scripts/2NF.sql>

<https://github.com/anibahs/MakeMyCareer/blob/main/FinalProject/Scripts/3NF.sql>