```
!pip install -q streamlit spacy PyMuPDF nltk pandas streamlit-option-menu
!python -m spacy download en_core_web_sm
```

```
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 44.3/44.3 kB 402.4 kB/s eta 0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 9.8/9.8 MB 22.7 MB/s eta 0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 20.0/20.0 MB 13.6 MB/s eta 0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 829.3/829.3 kB 14.7 MB/s eta 0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6.9/6.9 MB 21.4 MB/s eta 0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 79.1/79.1 kB 4.3 MB/s eta 0:00:00
Collecting en-core-web-sm==3.8.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.8.0/en_core_web_sm-3.8.0-py3-
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 12.8/12.8 MB 41.9 MB/s eta 0:00:00
✔ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
⚠ Restart to reload dependencies
If you are in a Jupyter or Colab notebook, you may need to restart Python in
order to load all the package's dependencies. You can do this by selecting the
'Restart kernel' or 'Restart runtime' option.
```

```
!mkdir -p data
```

```
%%writefile data/majors.csv
Major,Major_Category
GENERAL AGRICULTURE,Agriculture & Natural Resources
AGRICULTURE PRODUCTION AND MANAGEMENT,Agriculture & Natural Resources
AGRICULTURAL ECONOMICS,Agriculture & Natural Resources
ANIMAL SCIENCES,Agriculture & Natural Resources
FOOD SCIENCE,Agriculture & Natural Resources
PLANT SCIENCE AND AGRONOMY,Agriculture & Natural Resources
SOIL SCIENCE,Agriculture & Natural Resources
MISCELLANEOUS AGRICULTURE,Agriculture & Natural Resources
FORESTRY,Agriculture & Natural Resources
NATURAL RESOURCES MANAGEMENT,Agriculture & Natural Resources
FINE ARTS,Arts
DRAMA AND THEATER ARTS,Arts
MUSIC,Arts
VISUAL AND PERFORMING ARTS,Arts
COMMERCIAL ART AND GRAPHIC DESIGN,Arts
FILM VIDEO AND PHOTOGRAPHIC ARTS,Arts
STUDIO ARTS,Arts
MISCELLANEOUS FINE ARTS,Arts
ENVIRONMENTAL SCIENCE,Biology & Life Science
BIOLOGY,Biology & Life Science
BIOCHEMICAL SCIENCES,Biology & Life Science
BOTANY,Biology & Life Science
MOLECULAR BIOLOGY,Biology & Life Science
ECOLOGY,Biology & Life Science
GENETICS,Biology & Life Science
MICROBIOLOGY,Biology & Life Science
PHARMACOLOGY,Biology & Life Science
PHYSIOLOGY,Biology & Life Science
ZOOLOGY,Biology & Life Science
NEUROSCIENCE,Biology & Life Science
MISCELLANEOUS BIOLOGY,Biology & Life Science
COGNITIVE SCIENCE AND BIOPSYCHOLOGY,Biology & Life Science
GENERAL BUSINESS,Business
ACCOUNTING,Business
ACTUARIAL SCIENCE,Business
BUSINESS MANAGEMENT AND ADMINISTRATION,Business
OPERATIONS LOGISTICS AND E-COMMERCE,Business
BUSINESS ECONOMICS,Business
MARKETING AND MARKETING RESEARCH,Business
FINANCE,Business
HUMAN RESOURCES AND PERSONNEL MANAGEMENT,Business
INTERNATIONAL BUSINESS,Business
HOSPITALITY MANAGEMENT,Business
MANAGEMENT INFORMATION SYSTEMS AND STATISTICS,Business
MISCELLANEOUS BUSINESS & MEDICAL ADMINISTRATION,Business
COMMUNICATIONS,Communications & Journalism
JOURNALISM,Communications & Journalism
MASS MEDIA,Communications & Journalism
ADVERTISING AND PUBLIC RELATIONS,Communications & Journalism
COMMUNICATION TECHNOLOGIES,Computers & Mathematics
COMPUTER AND INFORMATION SYSTEMS,Computers & Mathematics
COMPUTER PROGRAMMING AND DATA PROCESSING,Computers & Mathematics
COMPUTER SCIENCE,Computers & Mathematics
Information Science and Technology,Computers & Mathematics
INFORMATION SCIENCES,Computers & Mathematics
COMPUTER ADMINISTRATION MANAGEMENT AND SECURITY,Computers & Mathematics
COMPUTER NETWORKING AND TELECOMMUNICATIONS,Computers & Mathematics
MATHEMATICS,Computers & Mathematics
```

```
APPLIED MATHEMATICS,Computers & Mathematics
STATISTICS AND DECISION SCIENCE,Computers & Mathematics
MATHEMATICS AND COMPUTER SCIENCE,Computers & Mathematics
GENERAL EDUCATION,Education
EDUCATIONAL ADMINISTRATION AND SUPERVISION,Education
SCHOOL STUDENT COUNSELING,Education
ELEMENTARY EDUCATION,Education
MATHEMATICS TEACHER EDUCATION,Education
PHYSICAL AND HEALTH EDUCATION TEACHING,Education
EARLY CHILDHOOD EDUCATION,Education
SCIENCE AND COMPUTER TEACHER EDUCATION,Education
SECONDARY TEACHER EDUCATION,Education
SPECIAL NEEDS EDUCATION,Education
SOCIAL SCIENCE OR HISTORY TEACHER EDUCATION,Education
TEACHER EDUCATION: MULTIPLE LEVELS,Education
LANGUAGE AND DRAMA EDUCATION,Education
ART AND MUSIC EDUCATION,Education
MISCELLANEOUS EDUCATION,Education
LIBRARY SCIENCE,Education
ARCHITECTURE,Engineering
GENERAL ENGINEERING,Engineering
AEROSPACE ENGINEERING,Engineering
BIOLOGICAL ENGINEERING,Engineering
ARCHITECTURAL ENGINEERING,Engineering
BIOMEDICAL ENGINEERING,Engineering
CHEMICAL ENGINEERING,Engineering
CIVIL ENGINEERING,Engineering
COMPUTER ENGINEERING,Engineering
ELECTRICAL ENGINEERING,Engineering
ENGINEERING MECHANICS PHYSICS AND SCIENCE,Engineering
ENVIRONMENTAL ENGINEERING,Engineering
GEOLOGICAL AND GEOPHYSICAL ENGINEERING,Engineering
INDUSTRIAL AND MANUFACTURING ENGINEERING,Engineering
MATERIALS ENGINEERING AND MATERIALS SCIENCE,Engineering
MECHANICAL ENGINEERING,Engineering
METALLURGICAL ENGINEERING,Engineering
MINING AND MINERAL ENGINEERING,Engineering
NAVAL ARCHITECTURE AND MARINE ENGINEERING,Engineering
NUCLEAR ENGINEERING,Engineering
PETROLEUM ENGINEERING,Engineering
MISCELLANEOUS ENGINEERING,Engineering
ENGINEERING TECHNOLOGIES,Engineering
ENGINEERING AND INDUSTRIAL MANAGEMENT,Engineering
ELECTRICAL ENGINEERING TECHNOLOGY,Engineering
INDUSTRIAL PRODUCTION TECHNOLOGIES,Engineering
MECHANICAL ENGINEERING RELATED TECHNOLOGIES,Engineering
MISCELLANEOUS ENGINEERING TECHNOLOGIES,Engineering
MATERIALS SCIENCE,Engineering
NUTRITION SCIENCES,Health
GENERAL MEDICAL AND HEALTH SERVICES,Health
COMMUNICATION DISORDERS SCIENCES AND SERVICES,Health
HEALTH AND MEDICAL ADMINISTRATIVE SERVICES,Health
MEDICAL ASSISTING SERVICES,Health
MEDICAL TECHNOLOGIES TECHNICIANS,Health
HEALTH AND MEDICAL PREPARATORY PROGRAMS,Health
NURSING,Health
PHARMACY PHARMACEUTICAL SCIENCES AND ADMINISTRATION,Health
TREATMENT THERAPY PROFESSIONS,Health
COMMUNITY AND PUBLIC HEALTH,Health
MISCELLANEOUS HEALTH MEDICAL PROFESSIONS,Health
AREA ETHNIC AND CIVILIZATION STUDIES,Humanities & Liberal Arts
LINGUISTICS AND COMPARATIVE LANGUAGE AND LITERATURE,Humanities & Liberal Arts
FRENCH GERMAN LATIN AND OTHER COMMON FOREIGN LANGUAGE STUDIES,Humanities & Liberal Arts
OTHER FOREIGN LANGUAGES,Humanities & Liberal Arts
ENGLISH LANGUAGE AND LITERATURE,Humanities & Liberal Arts
COMPOSITION AND RHETORIC,Humanities & Liberal Arts
LIBERAL ARTS,Humanities & Liberal Arts
HUMANITIES,Humanities & Liberal Arts
INTERCULTURAL AND INTERNATIONAL STUDIES,Humanities & Liberal Arts
PHILOSOPHY AND RELIGIOUS STUDIES,Humanities & Liberal Arts
THEOLOGY AND RELIGIOUS VOCATIONS,Humanities & Liberal Arts
ANTHROPOLOGY AND ARCHEOLOGY,Humanities & Liberal Arts
ART HISTORY AND CRITICISM,Humanities & Liberal Arts
HISTORY,Humanities & Liberal Arts
UNITED STATES HISTORY,Humanities & Liberal Arts
COSMETOLOGY SERVICES AND CULINARY ARTS,Industrial Arts & Consumer Services
FAMILY AND CONSUMER SCIENCES,Industrial Arts & Consumer Services
MILITARY TECHNOLOGIES,Industrial Arts & Consumer Services
PHYSICAL FITNESS PARKS RECREATION AND LEISURE,Industrial Arts & Consumer Services
CONSTRUCTION SERVICES,Industrial Arts & Consumer Services
ELECTRICAL MECHANICAL AND PRECISION TECHNOLOGIES AND PRODUCTION,Industrial Arts & Consumer Services
TRANSPORTATION SCIENCES AND TECHNOLOGIES,Industrial Arts & Consumer Services
```

```
    MULTI/INTERDISCIPLINARY STUDIES,Interdisciplinary
    COURT REPORTING,Law & Public Policy
    PRE-LAW AND LEGAL STUDIES,Law & Public Policy
    CRIMINAL JUSTICE AND FIRE PROTECTION,Law & Public Policy
    PUBLIC ADMINISTRATION,Law & Public Policy
    PUBLIC POLICY,Law & Public Policy
    LESS THAN BACHELORS DEGREE,Other
    PHYSICAL SCIENCES,Physical Sciences
    ASTRONOMY AND ASTROPHYSICS,Physical Sciences
    ATMOSPHERIC SCIENCES AND METEOROLOGY,Physical Sciences
    CHEMISTRY,Physical Sciences
    GEOLOGY AND EARTH SCIENCE,Physical Sciences
    GEOSCIENCES,Physical Sciences
    OCEANOGRAPHY,Physical Sciences
    PHYSICS,Physical Sciences
    MULTI-DISCIPLINARY OR GENERAL SCIENCE,Physical Sciences
    NUCLEAR INDUSTRIAL RADIOLOGY AND BIOLOGICAL TECHNOLOGIES,Physical Sciences
    PSYCHOLOGY,Psychology & Social Work
    EDUCATIONAL PSYCHOLOGY,Psychology & Social Work
    CLINICAL PSYCHOLOGY,Psychology & Social Work
    COUNSELING PSYCHOLOGY,Psychology & Social Work
    INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY,Psychology & Social Work
    SOCIAL PSYCHOLOGY,Psychology & Social Work
    MISCELLANEOUS PSYCHOLOGY,Psychology & Social Work
    HUMAN SERVICES AND COMMUNITY ORGANIZATION,Psychology & Social Work
    SOCIAL WORK,Psychology & Social Work
    INTERDISCIPLINARY SOCIAL SCIENCES,Social Science
    GENERAL SOCIAL SCIENCES,Social Science
    ECONOMICS,Social Science
    CRIMINOLOGY,Social Science
    GEOGRAPHY,Social Science
    INTERNATIONAL RELATIONS,Social Science
    POLITICAL SCIENCE AND GOVERNMENT,Social Science
    SOCIOLOGY,Social Science
    MISCELLANEOUS SOCIAL SCIENCES,Social Science
```

⇥ Writing data/majors.csv

```
%%writefile data/majors.csv
Major,Major_Category
GENERAL AGRICULTURE,Agriculture & Natural Resources
AGRICULTURE PRODUCTION AND MANAGEMENT,Agriculture & Natural Resources
AGRICULTURAL ECONOMICS,Agriculture & Natural Resources
ANIMAL SCIENCES,Agriculture & Natural Resources
FOOD SCIENCE,Agriculture & Natural Resources
PLANT SCIENCE AND AGRONOMY,Agriculture & Natural Resources
SOIL SCIENCE,Agriculture & Natural Resources
MISCELLANEOUS AGRICULTURE,Agriculture & Natural Resources
FORESTRY,Agriculture & Natural Resources
NATURAL RESOURCES MANAGEMENT,Agriculture & Natural Resources
FINE ARTS,Arts
DRAMA AND THEATER ARTS,Arts
MUSIC,Arts
VISUAL AND PERFORMING ARTS,Arts
COMMERCIAL ART AND GRAPHIC DESIGN,Arts
FILM VIDEO AND PHOTOGRAPHIC ARTS,Arts
STUDIO ARTS,Arts
MISCELLANEOUS FINE ARTS,Arts
ENVIRONMENTAL SCIENCE,Biology & Life Science
BIOLOGY,Biology & Life Science
BIOCHEMICAL SCIENCES,Biology & Life Science
BOTANY,Biology & Life Science
MOLECULAR BIOLOGY,Biology & Life Science
ECOLOGY,Biology & Life Science
GENETICS,Biology & Life Science
MICROBIOLOGY,Biology & Life Science
PHARMACOLOGY,Biology & Life Science
PHYSIOLOGY,Biology & Life Science
ZOOLOGY,Biology & Life Science
NEUROSCIENCE,Biology & Life Science
MISCELLANEOUS BIOLOGY,Biology & Life Science
COGNITIVE SCIENCE AND BIOPSYCHOLOGY,Biology & Life Science
GENERAL BUSINESS,Business
ACCOUNTING,Business
ACTUARIAL SCIENCE,Business
BUSINESS MANAGEMENT AND ADMINISTRATION,Business
OPERATIONS LOGISTICS AND E-COMMERCE,Business
BUSINESS ECONOMICS,Business
MARKETING AND MARKETING RESEARCH,Business
FINANCE,Business
HUMAN RESOURCES AND PERSONNEL MANAGEMENT,Business
INTERNATIONAL BUSINESS,Business
```

```
HOSPITALITY MANAGEMENT,Business
MANAGEMENT INFORMATION SYSTEMS AND STATISTICS,Business
MISCELLANEOUS BUSINESS & MEDICAL ADMINISTRATION,Business
COMMUNICATIONS,Communications & Journalism
JOURNALISM,Communications & Journalism
MASS MEDIA,Communications & Journalism
ADVERTISING AND PUBLIC RELATIONS,Communications & Journalism
COMMUNICATION TECHNOLOGIES,Computers & Mathematics
COMPUTER AND INFORMATION SYSTEMS,Computers & Mathematics
COMPUTER PROGRAMMING AND DATA PROCESSING,Computers & Mathematics
COMPUTER SCIENCE,Computers & Mathematics
Information Science and Technology,Computers & Mathematics
INFORMATION SCIENCES,Computers & Mathematics
COMPUTER ADMINISTRATION MANAGEMENT AND SECURITY,Computers & Mathematics
COMPUTER NETWORKING AND TELECOMMUNICATIONS,Computers & Mathematics
MATHEMATICS,Computers & Mathematics
APPLIED MATHEMATICS,Computers & Mathematics
STATISTICS AND DECISION SCIENCE,Computers & Mathematics
MATHEMATICS AND COMPUTER SCIENCE,Computers & Mathematics
GENERAL EDUCATION,Education
EDUCATIONAL ADMINISTRATION AND SUPERVISION,Education
SCHOOL STUDENT COUNSELING,Education
ELEMENTARY EDUCATION,Education
MATHEMATICS TEACHER EDUCATION,Education
PHYSICAL AND HEALTH EDUCATION TEACHING,Education
EARLY CHILDHOOD EDUCATION,Education
SCIENCE AND COMPUTER TEACHER EDUCATION,Education
SECONDARY TEACHER EDUCATION,Education
SPECIAL NEEDS EDUCATION,Education
SOCIAL SCIENCE OR HISTORY TEACHER EDUCATION,Education
TEACHER EDUCATION: MULTIPLE LEVELS,Education
LANGUAGE AND DRAMA EDUCATION,Education
ART AND MUSIC EDUCATION,Education
MISCELLANEOUS EDUCATION,Education
LIBRARY SCIENCE,Education
ARCHITECTURE,Engineering
GENERAL ENGINEERING,Engineering
AEROSPACE ENGINEERING,Engineering
BIOLOGICAL ENGINEERING,Engineering
ARCHITECTURAL ENGINEERING,Engineering
BIOMEDICAL ENGINEERING,Engineering
CHEMICAL ENGINEERING,Engineering
CIVIL ENGINEERING,Engineering
COMPUTER ENGINEERING,Engineering
ELECTRICAL ENGINEERING,Engineering
ENGINEERING MECHANICS PHYSICS AND SCIENCE,Engineering
ENVIRONMENTAL ENGINEERING,Engineering
GEOLOGICAL AND GEOPHYSICAL ENGINEERING,Engineering
INDUSTRIAL AND MANUFACTURING ENGINEERING,Engineering
MATERIALS ENGINEERING AND MATERIALS SCIENCE,Engineering
MECHANICAL ENGINEERING,Engineering
METALLURGICAL ENGINEERING,Engineering
MINING AND MINERAL ENGINEERING,Engineering
NAVAL ARCHITECTURE AND MARINE ENGINEERING,Engineering
NUCLEAR ENGINEERING,Engineering
PETROLEUM ENGINEERING,Engineering
MISCELLANEOUS ENGINEERING,Engineering
ENGINEERING TECHNOLOGIES,Engineering
ENGINEERING AND INDUSTRIAL MANAGEMENT,Engineering
ELECTRICAL ENGINEERING TECHNOLOGY,Engineering
INDUSTRIAL PRODUCTION TECHNOLOGIES,Engineering
MECHANICAL ENGINEERING RELATED TECHNOLOGIES,Engineering
MISCELLANEOUS ENGINEERING TECHNOLOGIES,Engineering
MATERIALS SCIENCE,Engineering
NUTRITION SCIENCES,Health
GENERAL MEDICAL AND HEALTH SERVICES,Health
COMMUNICATION DISORDERS SCIENCES AND SERVICES,Health
HEALTH AND MEDICAL ADMINISTRATIVE SERVICES,Health
MEDICAL ASSISTING SERVICES,Health
MEDICAL TECHNOLOGIES TECHNICIANS,Health
HEALTH AND MEDICAL PREPARATORY PROGRAMS,Health
NURSING,Health
PHARMACY PHARMACEUTICAL SCIENCES AND ADMINISTRATION,Health
TREATMENT THERAPY PROFESSIONS,Health
COMMUNITY AND PUBLIC HEALTH,Health
MISCELLANEOUS HEALTH MEDICAL PROFESSIONS,Health
AREA ETHNIC AND CIVILIZATION STUDIES,Humanities & Liberal Arts
LINGUISTICS AND COMPARATIVE LANGUAGE AND LITERATURE,Humanities & Liberal Arts
FRENCH GERMAN LATIN AND OTHER COMMON FOREIGN LANGUAGE STUDIES,Humanities & Liberal Arts
OTHER FOREIGN LANGUAGES,Humanities & Liberal Arts
ENGLISH LANGUAGE AND LITERATURE,Humanities & Liberal Arts
COMPOSITION AND RHETORIC,Humanities & Liberal Arts
```

```
LIBERAL ARTS,Humanities & Liberal Arts
HUMANITIES,Humanities & Liberal Arts
INTERCULTURAL AND INTERNATIONAL STUDIES,Humanities & Liberal Arts
PHILOSOPHY AND RELIGIOUS STUDIES,Humanities & Liberal Arts
THEOLOGY AND RELIGIOUS VOCATIONS,Humanities & Liberal Arts
ANTHROPOLOGY AND ARCHEOLOGY,Humanities & Liberal Arts
ART HISTORY AND CRITICISM,Humanities & Liberal Arts
HISTORY,Humanities & Liberal Arts
UNITED STATES HISTORY,Humanities & Liberal Arts
COSMETOLOGY SERVICES AND CULINARY ARTS,Industrial Arts & Consumer Services
FAMILY AND CONSUMER SCIENCES,Industrial Arts & Consumer Services
MILITARY TECHNOLOGIES,Industrial Arts & Consumer Services
PHYSICAL FITNESS PARKS RECREATION AND LEISURE,Industrial Arts & Consumer Services
CONSTRUCTION SERVICES,Industrial Arts & Consumer Services
"ELECTRICAL, MECHANICAL, AND PRECISION TECHNOLOGIES AND PRODUCTION",Industrial Arts & Consumer Services
TRANSPORTATION SCIENCES AND TECHNOLOGIES,Industrial Arts & Consumer Services
MULTI/INTERDISCIPLINARY STUDIES,Interdisciplinary
COURT REPORTING,Law & Public Policy
PRE-LAW AND LEGAL STUDIES,Law & Public Policy
CRIMINAL JUSTICE AND FIRE PROTECTION,Law & Public Policy
PUBLIC ADMINISTRATION,Law & Public Policy
PUBLIC POLICY,Law & Public Policy
N/A (less than bachelor's degree),
PHYSICAL SCIENCES,Physical Sciences
ASTRONOMY AND ASTROPHYSICS,Physical Sciences
ATMOSPHERIC SCIENCES AND METEOROLOGY,Physical Sciences
CHEMISTRY,Physical Sciences
GEOLOGY AND EARTH SCIENCE,Physical Sciences
GEOSCIENCES,Physical Sciences
OCEANOGRAPHY,Physical Sciences
PHYSICS,Physical Sciences
MULTI-DISCIPLINARY OR GENERAL SCIENCE,Physical Sciences
"NUCLEAR, INDUSTRIAL RADIOLOGY, AND BIOLOGICAL TECHNOLOGIES",Physical Sciences
PSYCHOLOGY,Psychology & Social Work
EDUCATIONAL PSYCHOLOGY,Psychology & Social Work
CLINICAL PSYCHOLOGY,Psychology & Social Work
COUNSELING PSYCHOLOGY,Psychology & Social Work
INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY,Psychology & Social Work
SOCIAL PSYCHOLOGY,Psychology & Social Work
MISCELLANEOUS PSYCHOLOGY,Psychology & Social Work
HUMAN SERVICES AND COMMUNITY ORGANIZATION,Psychology & Social Work
SOCIAL WORK,Psychology & Social Work
INTERDISCIPLINARY SOCIAL SCIENCES,Social Science
GENERAL SOCIAL SCIENCES,Social Science
ECONOMICS,Social Science
CRIMINOLOGY,Social Science
GEOGRAPHY,Social Science
INTERNATIONAL RELATIONS,Social Science
POLITICAL SCIENCE AND GOVERNMENT,Social Science
SOCIOLOGY,Social Science
MISCELLANEOUS SOCIAL SCIENCES,Social Science
```

⤓ Writing data/majors.csv

```
%%writefile data/position.csv
position,keywords
Software Engineer,"develop,design,code,program,debug"
Data Scientist,"research,analyze,data,statistical,experiment"
UX/UI Designer,"design,prototyping,usability,sketch,interface"
Product Manager,"strategy,roadmap,launch,market,analyze"
Network Engineer,"configure,administer,install,monitor,network"
Digital Marketing Specialist,"SEO,SEM,social media,analytics,campaign"
Financial Analyst,"analyze,forecast,budget,report,audit"
Human Resources Manager,"recruit,employee relations,compliance,training,benefits"
Content Writer,"write,edit,proofread,create,engage"
Graphic Designer,"illustrate,layout,typography,branding,visualize"
Sales Representative,"prospect,negotiate,pitch,close,client"
IT Support Specialist,"troubleshoot,resolve,install,configure,maintain"
Data Analyst,"analyze,interpret,report,statistics,query"
Systems Administrator,"administer,monitor,deploy,upgrade,configure"
Customer Support Specialist,"assist,troubleshoot,resolve,inquiries,escalation"
Supply Chain Manager,"optimize,coordinate,logistics,inventory,procure"
Quality Assurance Engineer,"test,verify,validate,automate,debug"
DevOps Engineer,"deploy,automate,configure,monitor,orchestrate"
UI/UX Developer,"design,front-end,usability,implement,interact"
Technical Writer,"document,create,edit,review,revise"
Cybersecurity Analyst,"monitor,protect,analyze,secure,detect"
Research Scientist,"conduct,experiment,analyze,observe,discover"
Operations Manager,"optimize,coordinate,efficiency,supervise,process"
Mechanical Engineer,"design,analyze,prototype,manufacture,validate"
Architect,"design,plan,draw,create,construct"
Healthcare Administrator,"manage,coordinate,regulate,plan,administer"
```

```
Business Analyst,"analyze,document,requirements,evaluate,improve"
Electrical Engineer,"design,analyze,implement,test,develop"
UX Researcher,"plan,conduct,analyze,usability,observe"
Project Manager,"plan,execute,coordinate,communicate,monitor"
Legal Counsel,"advise,review,negotiate,legal research,draft"
Data Engineer,"develop,build,maintain,design,implement"
Biomedical Engineer,"design,develop,implement,validate,test"
Social Media Manager,"content,engage,promote,analytics,monitor"
Systems Analyst,"analyze,design,implement,evaluate,modify"
Chemical Engineer,"design,develop,implement,process,analyze"
AI/ML Engineer,"develop,train,implement,analyze,optimize"
Technical Support Engineer,"troubleshoot,resolve,install,configure,assist"
Pharmacist,"dispense,review,monitor,consult,administer"
Interior Designer,"design,space,plan,coordinate,visualize"
Event Planner,"organize,plan,coordinate,execute,manage"
Civil Engineer,"design,analyze,construct,inspect,manage"
Environmental Scientist,"analyze,study,assess,monitor,evaluate"
Digital Analyst,"analyze,interpret,report,web,metrics"
UX/UI Architect,"design,prototype,usability,lead,implement"
Chief Financial Officer,"oversee,manage,financial strategy,report,analyze"
Chief Technology Officer,"lead,innovate,technology strategy,develop,execute"
Chief Marketing Officer,"lead,strategize,execute,analyze,promote"
Chief Executive Officer,"lead,manage,strategize,execute,oversee"
```

⤓ Writing data/position.csv

```
%%writefile data/feedback_data.csv
User Name: deep
Feedback: very good
Timestamp: 2023-11-27 14:17:30
--------------------------------------------------
User Name: dev
Feedback: ddd
Timestamp: 2023-11-27 14:18:25
--------------------------------------------------
```

⤓ Writing data/feedback_data.csv

```
%%writefile data/newSkills.csv
Python
Java
JavaScript
HTML
CSS
SQL
React
Angular
Node.js
MongoDB
AWS
Azure
Docker
Kubernetes
Git
Machine Learning
Data Analysis
Data Visualization
Tableau
Power BI
R
MATLAB
C++
C#
Go
Ruby
PHP
Swift
Rust
Dart
Kotlin
NoSQL
Scala
Perl
Haskell
Bash
Shell
Linux
Windows
MacOS
Android
iOS
```

```
⇥  Writing data/newSkills.csv
```

```
%%writefile data/sugestedSkills.csv
software engineer,Python,Java,JavaScript,React,Django,Git
software developer,Programming Basics,Software Development Lifecycle Understanding,Version Control Systems,Code Debugging
data scientist,Python,R,Machine Learning,Statistics,SQL
graphic designer,Adobe Photoshop,Illustrator,UI/UX Design,Typography
marketing analyst,Market Research,Data Analysis,Marketing Strategy,Excel,SQL
financial analyst,Financial Modeling,Accounting,Excel,Financial Analysis
frontend developer,HTML,CSS,JavaScript,React,Angular,Bootstrap
backend developer,Python,Java,C++,Node.js,SQL,MongoDB
network engineer,Networking Protocols,Cisco,Firewalls,Security,LAN/WAN
UX/UI designer,Wireframing,Prototyping,User Research,Adobe XD,Sketch
project manager,Leadership,Project Planning,Team Management,Communication
business analyst,Business Intelligence,Requirements Gathering,SQL,Tableau
digital marketer,SEO,SEM,Social Media Marketing,Google Analytics,Content Strategy
mobile app developer,Swift,Java,Kotlin,Android Studio,iOS Development
AI engineer,Machine Learning,Deep Learning,Neural Networks,Python,TensorFlow
cloud architect,Amazon Web Services (AWS),Microsoft Azure,GCP,Docker,Kubernetes
cybersecurity analyst,Information Security,Penetration Testing,Incident Response
IT support specialist,Troubleshooting,Help Desk Support,Hardware Installation,Network Administration
data engineer,ETL,Data Warehousing,Apache Spark,Hadoop,SQL
full-stack developer,HTML,CSS,JavaScript,Python,Node.js,React
UI/UX researcher,User Testing,Qualitative Research,Data Analysis,Usability Testing
product manager,Product Development,Market Analysis,Strategic Planning,Agile Methodologies
systems administrator,Server Administration,Virtualization,Scripting,Active Directory
```

```
⇥  Writing data/sugestedSkills.csv
```

```
%%writefile data/UpdatedSkills.csv
Python
Java
JavaScript
React
Angular
Node.js
MongoDB
SQL
AWS
Azure
Docker
Kubernetes
Git
Machine Learning
Data Analysis
```

```
⇥  Writing data/UpdatedSkills.csv
```

```python
%%writefile colab_app.py
import streamlit as st
import spacy
import fitz  # PyMuPDF
import sqlite3
import csv
import nltk
import re
from datetime import datetime
import base64
import pandas as pd
from spacy.matcher import Matcher
import os

# Download required NLTK data
nltk.download('punkt')

# Load the spaCy model
nlp = spacy.load('en_core_web_sm')

# Function to create a table for PDFs in SQLite database if it doesn't exist
def create_table():
    conn = sqlite3.connect('data/user_pdfs.db')
    cursor = conn.cursor()
    cursor.execute('''
        CREATE TABLE IF NOT EXISTS user_uploaded_pdfs (
            id INTEGER PRIMARY KEY,
            name TEXT NOT NULL,
            data BLOB NOT NULL
        )
    ''')
    conn.commit()
    conn.close()
```

```
                  commit()

    # Function to insert PDF into the SQLite database
    def insert_pdf(name, data):
        conn = sqlite3.connect('data/user_pdfs.db')
        cursor = conn.cursor()
        cursor.execute('INSERT INTO user_uploaded_pdfs (name, data) VALUES (?, ?)', (name, data))
        conn.commit()
        conn.close()

    # Function to load keywords from CSV
    def load_keywords(file_path):
        try:
            with open(file_path, 'r') as file:
                reader = csv.reader(file)
                return set(row[0] for row in reader)
        except:
            return set()

    # Function to extract text from PDF
    def extract_text_from_pdf(file):
        pdf_document = fitz.open(stream=file.read(), filetype="pdf")
        text = ""
        for page_num in range(len(pdf_document)):
            page = pdf_document.load_page(page_num)
            text += page.get_text()
        return text

    # Function to extract name from document
    def extract_name(doc):
        for ent in doc.ents:
            if ent.label_ == 'PERSON':
                names = ent.text.split()
                if len(names) >= 2 and names[0].istitle() and names[1].istitle():
                    return names[0], ' '.join(names[1:])
        return "", ""

    # Function to extract email from document
    def extract_email(doc):
        matcher = Matcher(nlp.vocab)
        email_pattern = [{'LIKE_EMAIL': True}]
        matcher.add('EMAIL', [email_pattern])

        matches = matcher(doc)
        for match_id, start, end in matches:
            if match_id == nlp.vocab.strings['EMAIL']:
                return doc[start:end].text
        return ""

    # Function to extract contact number from resume
    def extract_contact_number_from_resume(doc):
        contact_number = None
        text = doc.text
        pattern = r"\b(?:\+?\d{1,3}[-.\s]?)?\(?\d{3}\)?[-.\s]?\d{3}[-.\s]?\d{4}\b"
        match = re.search(pattern, text)
        if match:
            contact_number = match.group()
        return contact_number

    # Function to extract education from resume
    def extract_education_from_resume(doc):
        universities = []
        for entity in doc.ents:
            if entity.label_ == "ORG" and ("university" in entity.text.lower() or "college" in entity.text.lower() or "institute"
                universities.append(entity.text)
        return universities

    # Function to extract skills from document
    def extract_skills(doc):
        skills_csv = set()
        skills_ner = set()

        # Extract skills from CSV
        try:
            skills_keywords = load_keywords('data/newSkills.csv')
            for keyword in skills_keywords:
                if keyword.lower() in doc.text.lower():
                    skills_csv.add(keyword)
        except:
            pass

        # Extract skills using NER
        for token in doc:
            if token.pos_ == 'NOUN' and token.text.isalpha() and len(token.text) > 1:
```

```python
        if token.pos_ == "NOUN" and token.text.isalpha() and len(token.text) > 1:
            skills_ner.add(token.text.lower())

    # Combine skills
    combined_skills = skills_csv.union(skills_ner)
    return list(combined_skills)

# Function to extract major from document
def extract_major(doc):
    try:
        major_keywords = load_keywords('data/majors.csv')
        for keyword in major_keywords:
            if keyword.lower() in doc.text.lower():
                return keyword
    except:
        pass
    return ""

# Function to extract experience from document
def extract_experience(doc):
    verbs = [token.text for token in doc if token.pos_ == 'VERB']

    senior_keywords = ['lead', 'manage', 'direct', 'oversee', 'supervise', 'orchestrate', 'govern']
    mid_senior_keywords = ['develop', 'design', 'analyze', 'implement', 'coordinate', 'execute', 'strategize']
    mid_junior_keywords = ['assist', 'support', 'collaborate', 'participate', 'aid', 'facilitate', 'contribute']

    if any(keyword in verbs for keyword in senior_keywords):
        level_of_experience = "Senior"
    elif any(keyword in verbs for keyword in mid_senior_keywords):
        level_of_experience = "Mid-Senior"
    elif any(keyword in verbs for keyword in mid_junior_keywords):
        level_of_experience = "Mid-Junior"
    else:
        level_of_experience = "Entry Level"

    return {
        'level_of_experience': level_of_experience,
        'suggested_position': "Position Not Identified"
    }

# Function to calculate resume score
def calculate_resume_score(resume_info):
    score = 0
    if resume_info['first_name'] and resume_info['last_name']:
        score += 25
    if resume_info['email']:
        score += 25
    if resume_info['degree_major']:
        score += 25
    if resume_info['skills']:
        score += 25
    return score

# Function to suggest skills for job
def suggest_skills_for_job(desired_job):
    try:
        job_skills_mapping = {}
        with open('data/sugestedSkills.csv', newline='') as csvfile:
            reader = csv.reader(csvfile)
            for row in reader:
                job_title = row[0].lower()
                skills = row[1:]
                job_skills_mapping[job_title] = skills

        desired_job_lower = desired_job.lower()
        if desired_job_lower in job_skills_mapping:
            return job_skills_mapping[desired_job_lower]
    except:
        pass
    return []

# Function to process user mode
def process_user_mode():
    create_table()

    st.title("Resume Parser using NLP")
    uploaded_file = st.file_uploader("Upload a PDF resume", type="pdf")

    if uploaded_file:
        st.write("File uploaded successfully!")

        pdf_name = uploaded_file.name
        pdf_data = uploaded_file.getvalue()
```

```python
        # Insert the uploaded PDF into the database
        insert_pdf(pdf_name, pdf_data)

        # Extract text from PDF
        text = extract_text_from_pdf(uploaded_file)
        doc = nlp(text)

        # Extract resume information
        first_name, last_name = extract_name(doc)
        email = extract_email(doc)
        contact_number = extract_contact_number_from_resume(doc)
        skills = extract_skills(doc)
        degree_major = extract_major(doc)
        experience = extract_experience(doc)

        # Display extracted information
        st.markdown('<hr>', unsafe_allow_html=True)
        st.header("Extracted Information:")
        st.write(f"First Name: {first_name}")
        st.write(f"Last Name: {last_name}")
        st.write(f"Email: {email}")
        st.write(f"Phone Number: {contact_number}")
        st.write(f"Degree/Major: {degree_major}")

        st.markdown('<hr>', unsafe_allow_html=True)
        st.header("Education:")
        education_info = extract_education_from_resume(doc)
        st.write(', '.join(education_info) if education_info else "No education information found")

        st.markdown('<hr>', unsafe_allow_html=True)
        st.header("Skills:")
        st.write(', '.join(skills))

        st.markdown('<hr>', unsafe_allow_html=True)
        st.header("Experience:")
        st.write(f"Level of Experience: {experience['level_of_experience']}")
        st.write(f"Suggested Position: {experience['suggested_position']}")

        # Calculate and display resume score
        resume_info = {
            'first_name': first_name,
            'last_name': last_name,
            'email': email,
            'degree_major': degree_major,
            'skills': skills
        }
        resume_score = calculate_resume_score(resume_info)

        st.markdown('<hr>', unsafe_allow_html=True)
        st.header("Resume Score:")
        st.write(f"**Resume Score:** {resume_score}/100")

        # Display a custom-styled progress bar with gradient colors
        percentage = resume_score
        percentage_str = str(percentage)
        bar = (
            f'<div style="background: linear-gradient(90deg, #f63366 {percentage_str}%, #d6d6d6 {percentage_str}%);'
            'height: 30px; border-radius: 5px; display: flex; align-items: center;">'
            f'<div style="color: white; text-align: center; width: 100%;">{percentage}%</div>'
            '</div>'
        )
        st.markdown(bar, unsafe_allow_html=True)

        # Suggest skills for desired job
        st.markdown('<hr>', unsafe_allow_html=True)
        st.header("Suggested Skills for the Desired Job:")
        desired_job = st.text_input("Enter the job you are looking for:")
        suggested_skills = suggest_skills_for_job(desired_job)
        st.write(suggested_skills)

# Function to process recruiters mode
def process_recruiters_mode():
    st.title("Recruiter's Panel")

    # File upload for resumes
    uploaded_files = st.file_uploader("Upload resumes (PDF)", accept_multiple_files=True)

    # Input for required skills
    required_skills_input = st.text_input("Enter required skills (comma-separated)", "")
    required_skills = [skill.strip().lower() for skill in required_skills_input.split(',') if skill.strip()]

    # Button to save required skills to UpdatedSkills.csv
```

```python
        if st.button("Save Required Skills"):
            with open('data/UpdatedSkills.csv', 'a', newline='') as file:
                writer = csv.writer(file)
                for skill in required_skills:
                    writer.writerow([skill])

        # Check skills and extract names in uploaded resumes and collect all skills found
        all_skills_found = set()
        if uploaded_files:
            for file in uploaded_files:
                text = extract_text_from_pdf(file)
                doc = nlp(text)

                # Extract candidate name
                candidate_name = ""
                for ent in doc.ents:
                    if ent.label_ == 'PERSON':
                        candidate_name = ent.text
                        break

                if not candidate_name:
                    candidate_name = "Candidate name not found"

                # Display candidate information
                st.write("-" * 30)
                st.write(f"**File Name:** {file.name}")
                st.subheader(f"**Candidate Name:** {candidate_name}")

                # Extract and display all skills
                all_skills = set()
                for token in doc:
                    if token.pos_ == 'NOUN' and token.text.isalpha() and len(token.text) > 1:
                        all_skills.add(token.text.lower())

                if all_skills:
                    st.subheader("\n**All Skills Parsed from Resume:**")
                    st.write(", ".join(all_skills))
                else:
                    st.subheader("\n**No Skills Parsed from Resume**")

                # Check for required skills
                skills_found = set()
                for skill in required_skills:
                    if skill in all_skills:
                        skills_found.add(skill)

                st.subheader("\n**Skills Found or Not Found:**\n")
                for skill in required_skills:
                    if skill in skills_found:
                        st.write(f"- {skill}: Found")
                    else:
                        st.write(f"- {skill}: Not Found")

                all_skills_found.update(skills_found)

# Function to process feedback mode
def process_feedback_mode():
    st.title("Feedback Section")
    st.subheader("Provide Feedback")

    # Feedback Form
    user_name = st.text_input("Your Name:")
    feedback = st.text_area("Provide feedback on the resume parser:", height=100)
    if st.button("Submit Feedback"):
        timestamp = datetime.now().strftime("%Y-%m-%d %H:%M:%S")
        with open('data/feedback_data.csv', 'a') as file:
            file.write(f"User Name: {user_name}\n")
            file.write(f"Feedback: {feedback}\n")
            file.write(f"Timestamp: {timestamp}\n")
            file.write("-" * 50 + "\n")
        st.success("Feedback submitted successfully!")

# Function to process admin mode
def process_admin_mode():
    st.title("Admin Panel")
    st.subheader("Authentication Required")

    # Admin authentication
    username = st.text_input("Username:")
    password = st.text_input("Password:", type="password")
    if st.button("Login"):
        if username == "deep" and password == "dp10":
            st.success("Authentication successful!")
```

```
                st.success("Authentication successful.")

            # Display uploaded PDFs
            try:
                conn = sqlite3.connect('data/user_pdfs.db')
                cursor = conn.cursor()
                cursor.execute("SELECT id, name FROM user_uploaded_pdfs")
                uploaded_pdfs = cursor.fetchall()
                conn.close()

                if uploaded_pdfs:
                    st.subheader("Uploaded Resumes")

                    pdf_data_list = []
                    for pdf_id, pdf_name in uploaded_pdfs:
                        try:
                            conn = sqlite3.connect('data/user_pdfs.db')
                            cursor = conn.cursor()
                            cursor.execute("SELECT name, data FROM user_uploaded_pdfs WHERE id=?", (pdf_id,))
                            pdf_data = cursor.fetchone()
                            conn.close()

                            if pdf_data:
                                pdf_b64 = base64.b64encode(pdf_data[1]).decode('utf-8')
                                href = f'<a href="data:application/pdf;base64,{pdf_b64}" download="{pdf_name}">Download</a>'
                                pdf_data_list.append({"ID": pdf_id, "Name": pdf_name, "Download (Resume)": href})
                            else:
                                st.warning(f"Error retrieving PDF data for ID: {pdf_id}")
                        except sqlite3.Error as e:
                            st.error(f"Error fetching PDF data: {e}")

                    # Create DataFrame from the list of dictionaries
                    pdf_table = pd.DataFrame(pdf_data_list)
                    # Display the table with HTML links using markdown
                    st.markdown(pdf_table.to_html(escape=False), unsafe_allow_html=True)
                else:
                    st.warning("No uploaded PDFs available.")
            except sqlite3.Error as e:
                st.error(f"Error fetching uploaded PDFs: {e}")

            # Display feedback data
            st.markdown('---')
            try:
                feedback_data = pd.read_csv('data/feedback_data.csv')
                latest_feedback = feedback_data.tail(10)  # Fetching latest 10 feedbacks

                st.subheader("Latest Feedbacks")
                st.write(latest_feedback)  # Display latest 10 feedbacks

                if st.button("View More Feedbacks"):
                    st.write(feedback_data)  # Display all feedbacks if requested by admin
            except FileNotFoundError:
                st.warning("No feedback data available.")
        else:
            st.error("Authentication failed. Please try again.")

# Main function
def main():
    st.set_page_config(page_title="Resume Parser", page_icon="✅")

    # Sidebar
    st.sidebar.title("Navigation")
    app_mode = st.sidebar.selectbox("Choose an option", ["Users", "Recruiters", "Feedback", "Admin"])

    if app_mode == "Users":
        process_user_mode()
    elif app_mode == "Recruiters":
        process_recruiters_mode()
    elif app_mode == "Admin":
        process_admin_mode()
    elif app_mode == "Feedback":
        process_feedback_mode()

if __name__ == "__main__":
    main()
```

⤓ Writing colab_app.py

```
from google.colab.output import eval_js
print(eval_js("google.colab.kernel.proxyPort(8501)"))
```

```
# Run the Streamlit app
!streamlit run colab_app.py --server.port 8501 --server.address 0.0.0.0 --server.headless True --server.enableCORS False --s
```

••• https://8501-m-s-1ud6oycgebze5-b.us-west3-1.prod.colab.dev

Collecting usage statistics. To deactivate, set browser.gatherUsageStats to false.

  You can now view your Streamlit app in your browser.

  URL: **http://0.0.0.0:8501**

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
2025-04-10 03:11:42.564 Examining the path of torch.classes raised:
Traceback (most recent call last):
  File "/usr/local/lib/python3.11/dist-packages/streamlit/web/bootstrap.py", line 347, in run
    if asyncio.get_running_loop().is_running():
       ^^^^^^^^^^^^^^^^^^^^^^^^^^^
RuntimeError: no running event loop

During handling of the above exception, another exception occurred:

Traceback (most recent call last):
  File "/usr/local/lib/python3.11/dist-packages/streamlit/watcher/local_sources_watcher.py", line 217, in get_module_pat
    potential_paths = extract_paths(module)
                      ^^^^^^^^^^^^^^^^^^^^^
  File "/usr/local/lib/python3.11/dist-packages/streamlit/watcher/local_sources_watcher.py", line 210, in <lambda>
    lambda m: list(m.__path__._path),
                   ^^^^^^^^^^^^^^^
  File "/usr/local/lib/python3.11/dist-packages/torch/_classes.py", line 13, in __getattr__
    proxy = torch._C._get_custom_class_python_wrapper(self.name, attr)
            ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
RuntimeError: Tried to instantiate class '__path__._path', but it does not exist! Ensure that it is registered via torch
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

Start coding or generate with AI.