

```
get_ipython().system('pip install streamlit')
get_ipython().system('pip install pandas')
get_ipython().system('pip install spacy')
get_ipython().system('pip install nltk')
get_ipython().system('pip install PyPDF2')
get_ipython().system('pip install scikit-learn')
```

Collecting streamlit

Downloading streamlit-1.51.0-py3-none-any.whl.metadata (9.5 kB)
Requirement already satisfied: altair!=5.4.0,!5.4.1,<6,>=4.0 in /usr/local/lib/python3.12/dist-packages (from streamlit)
Requirement already satisfied: blinker<2,>=1.5.0 in /usr/local/lib/python3.12/dist-packages (from streamlit) (1.9.0)
Requirement already satisfied: cachetools<7,>=4.0 in /usr/local/lib/python3.12/dist-packages (from streamlit) (5.5.2)
Requirement already satisfied: click<9,>=7.0 in /usr/local/lib/python3.12/dist-packages (from streamlit) (8.3.1)
Requirement already satisfied: numpy<3,>=1.23 in /usr/local/lib/python3.12/dist-packages (from streamlit) (2.0.2)
Requirement already satisfied: packaging<26,>=20 in /usr/local/lib/python3.12/dist-packages (from streamlit) (25.0)
Requirement already satisfied: pandas<3,>=1.4.0 in /usr/local/lib/python3.12/dist-packages (from streamlit) (2.2.2)
Requirement already satisfied: pillow<13,>=7.1.0 in /usr/local/lib/python3.12/dist-packages (from streamlit) (11.3.0)
Requirement already satisfied: protobuf<7,>=3.20 in /usr/local/lib/python3.12/dist-packages (from streamlit) (5.29.5)
Requirement already satisfied: pyarrow<22,>=7.0 in /usr/local/lib/python3.12/dist-packages (from streamlit) (18.1.0)
Requirement already satisfied: requests<3,>=2.27 in /usr/local/lib/python3.12/dist-packages (from streamlit) (2.32.4)
Requirement already satisfied: tenacity<10,>=8.1.0 in /usr/local/lib/python3.12/dist-packages (from streamlit) (8.5.0)
Requirement already satisfied: toml<2,>=0.10.1 in /usr/local/lib/python3.12/dist-packages (from streamlit) (0.10.2)
Requirement already satisfied: typing-extensions<5,>=4.4.0 in /usr/local/lib/python3.12/dist-packages (from streamlit) (4.12.0)
Requirement already satisfied: watchdog<7,>=2.1.5 in /usr/local/lib/python3.12/dist-packages (from streamlit) (6.0.0)
Requirement already satisfied: gitpython!=3.1.19,<4,>=3.0.7 in /usr/local/lib/python3.12/dist-packages (from streamlit) (3.1.40)
Collecting pydeck<1,>=0.8.0b4 (from streamlit)

Downloading pydeck-0.9.1-py2.py3-none-any.whl.metadata (4.1 kB)
Requirement already satisfied: tornado!=6.5.0,<7,>=6.0.3 in /usr/local/lib/python3.12/dist-packages (from streamlit) (6.5.0)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.12/dist-packages (from altair!=5.4.0,!5.4.1,<6,>=4.0->streamlit) (3.1.4)
Requirement already satisfied: jsonschema<3.0 in /usr/local/lib/python3.12/dist-packages (from altair!=5.4.0,!5.4.1,<6,>=4.0->streamlit) (4.17.0)
Requirement already satisfied: narwhals>=1.14.2 in /usr/local/lib/python3.12/dist-packages (from altair!=5.4.0,!5.4.1,<6,>=4.0->streamlit) (1.42.1)
Requirement already satisfied: gitdb<5,>=4.0.1 in /usr/local/lib/python3.12/dist-packages (from gitpython!=3.1.19,<4,>=3.0.7->streamlit) (4.0.10)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas<3,>=1.4.0->streamlit) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas<3,>=1.4.0->streamlit) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas<3,>=1.4.0->streamlit) (2025.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.27->streamlit) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.27->streamlit) (3.10.1)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.27->streamlit) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.27->streamlit) (2025.11.12)
Requirement already satisfied: smmap<6,>=3.0.1 in /usr/local/lib/python3.12/dist-packages (from gitdb<5,>=4.0.1->gitpython) (5.0.6)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from Jinja2->altair!=5.4.0,!5.4.1,<6,>=4.0->streamlit) (3.0.2)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.12/dist-packages (from jsonschema>=3.0->altair!=5.4.0,!5.4.1,<6,>=4.0->streamlit) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.12/dist-packages (from jsonschema>=3.0->altair!=5.4.0,!5.4.1,<6,>=4.0->streamlit) (2025.9.1)
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.12/dist-packages (from jsonschema>=3.0->altair!=5.4.0,!5.4.1,<6,>=4.0->streamlit) (0.36.0)
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.12/dist-packages (from jsonschema>=3.0->altair!=5.4.0,!5.4.1,<6,>=4.0->streamlit) (0.23.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas<3,>=1.4.0->streamlit) (1.17.0)
Downloading streamlit-1.51.0-py3-none-any.whl (10.2 MB)

10.2/10.2 MB 100.3 MB/s eta 0:00:00

Downloading pydeck-0.9.1-py2.py3-none-any.whl (6.9 MB)

6.9/6.9 MB 108.1 MB/s eta 0:00:00

Installing collected packages: pydeck, streamlit

Successfully installed pydeck-0.9.1 streamlit-1.51.0

Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages (2.2.2)
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Requirement already satisfied: spacy in /usr/local/lib/python3.12/dist-packages (3.8.11)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.0.15)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.13)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.0.12)

```
!python -m spacy download en_core_web_sm
```

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-none-any.whl 12.8/12.8 MB 108.5 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.

```
from PyPDF2 import PdfReader
```

```
def extract_text_from_pdf(pdf_file):
    text = ""
    reader = PdfReader(pdf_file)
    for page in reader.pages:
        extracted = page.extract_text()
```

```

        if extracted:
            text += extracted
        return text

def extract_text_from_text_input(text_input):
    return text_input.strip()

```

```

input_text = input("Paste your resume text here:\n")
resume_text = extract_text_from_text_input(input_text)
print(resume_text[:800])

```

```

-----
KeyboardInterrupt                                Traceback (most recent call last)
/tmp/ipython-input-204784162.py in <cell line: 0>()
----> 1 input_text = input("Paste your resume text here:\n")
      2 resume_text = extract_text_from_text_input(input_text)
      3 print(resume_text[:800])

----- 1 frames -----
/usr/local/lib/python3.12/dist-packages/ipykernel/kernelbase.py in _input_request(self, prompt, ident, parent, password)
    1217         except KeyboardInterrupt:
    1218             # re-raise KeyboardInterrupt, to truncate traceback
-> 1219             raise KeyboardInterrupt("Interrupted by user") from None
    1220     except Exception:
    1221         self.log.warning("Invalid Message:", exc_info=True)

KeyboardInterrupt: Interrupted by user

```

```

# Test 1: PDF Upload
from google.colab import files
uploaded = files.upload()

file_name = list(uploaded.keys())[0]
with open(file_name, "rb") as pdf:
    resume_text = extract_text_from_pdf(pdf)

print(resume_text[:800])

```

Choose Files No file chosen

```

-----
IndexError                                Traceback (most recent call last)
/tmp/ipython-input-2828919047.py in <cell line: 0>()
      3 uploaded = files.upload()
      4
----> 5 file_name = list(uploaded.keys())[0]
      6 with open(file_name, "rb") as pdf:
      7     resume_text = extract_text_from_pdf(pdf)

IndexError: list index out of range

```

Next steps: [Explain error](#)

```

# --- Create jobs.csv, load it, define matching, and test ---
import pandas as pd
from io import StringIO
import os

csv_content = """Job Role,Required Skills,Job Description,Salary Range,Experience
Data Analyst,"excel, sql, power bi, python","Analyze datasets and generate insights","4-8 LPA","0-2 years"
Python Developer,"python, django, flask, sql","Build backend APIs and automation tools","5-9 LPA","0-2 years"
ML Engineer,"python, machine learning, deep learning, statistics","Train and deploy ML models","7-12 LPA","0-2 years"
Web Developer,"html, css, javascript, react, node","Develop front-end and full-stack web applications","4-7 LPA","0-2 years"
Business Analyst,"excel, communication, problem solving","Identify business issues and prepare reports","4-7 LPA","0-2 years"
HR Executive,"communication, ms office, recruitment","Manage hiring and employee onboarding","3-6 LPA","0-2 years"
Graphic Designer,"canva, photoshop, creativity","Create posters, logos, and marketing design","3-6 LPA","0-2 years"
Content Writer,"content writing, creativity, ms word","Write blogs, articles, reports, and scripts","2-5 LPA","0-2 years"
Digital Marketer,"social media, communication, canva","Promote brands through online campaigns","3-6 LPA","0-2 years"
"""

# Save the CSV file to the notebook working directory
with open("jobs.csv", "w", encoding="utf-8") as f:
    f.write(csv_content)

# Confirm file exists
print("jobs.csv created:", os.path.exists("jobs.csv"))

```

```

# Load into DataFrame
jobs_dataset = pd.read_csv("jobs.csv")
print("\n--- jobs_dataset head ---")
print(jobs_dataset.head())

# Define match_jobs (same logic as before)
def match_jobs(candidate_skills, jobs_df=jobs_dataset):
    results = []
    cand = set([s.lower().strip() for s in candidate_skills])
    for _, row in jobs_df.iterrows():
        job_skills = [s.strip().lower() for s in str(row['Required Skills']).split(",")]
        matched = cand & set(job_skills)
        score = (len(matched) / len(job_skills)) * 100 if len(job_skills) > 0 else 0
        results.append({
            "Job Role": row['Job Role'],
            "Match Score (%)": round(score, 2),
            "Skills Matched": sorted(list(matched)),
            "Skills Missing": sorted(list(set(job_skills) - cand)),
            "Salary Range": row['Salary Range'],
            "Experience": row['Experience'],
            "Job Description": row['Job Description']
        })
    results = sorted(results, key=lambda x: x["Match Score (%)"], reverse=True)
    return results

# --- TEST ---
# If you already have `skills_found` from your extractor, use it. Otherwise, use this sample:
sample_skills = ["python", "excel", "power bi", "communication"] # replace with your skills_found
print("\n--- Testing with sample_skills:", sample_skills, "---")
results = match_jobs(sample_skills)
for r in results[:5]:
    print("\nJob:", r["Job Role"])
    print("Score:", r["Match Score (%)"], "%")
    print("Matched:", r["Skills Matched"])
    print("Missing:", r["Skills Missing"])

```

jobs.csv created: True

```

--- jobs_dataset head ---

```

	Job Role	Required Skills
0	Data Analyst	excel, sql, power bi, python
1	Python Developer	python, django, flask, sql
2	ML Engineer	python, machine learning, deep learning, stati...
3	Web Developer	html, css, javascript, react, node
4	Business Analyst	excel, communication, problem solving

	Job Description	Salary Range	Experience
0	Analyze datasets and generate insights	4-8 LPA	0-2 years
1	Build backend APIs and automation tools	5-9 LPA	0-2 years
2	Train and deploy ML models	7-12 LPA	0-2 years
3	Develop front-end and full-stack web applications	4-7 LPA	0-2 years
4	Identify business issues and prepare reports	4-7 LPA	0-2 years

```

--- Testing with sample_skills: ['python', 'excel', 'power bi', 'communication'] ---

Job: Data Analyst
Score: 75.0 %
Matched: ['excel', 'power bi', 'python']
Missing: ['sql']

Job: Business Analyst
Score: 66.67 %
Matched: ['communication', 'excel']
Missing: ['problem solving']

Job: HR Executive
Score: 33.33 %
Matched: ['communication']
Missing: ['ms office', 'recruitment']

Job: Digital Marketer
Score: 33.33 %
Matched: ['communication']
Missing: ['canva', 'social media']

Job: Python Developer
Score: 25.0 %
Matched: ['python']
Missing: ['django', 'flask', 'sql']

```

