```
!pip install sentence-transformers
```

```
Requirement already satisfied: sentence-transformers in /usr/local/lib/python3.12/dist-packages (5.2.2)
Requirement already satisfied: transformers<6.0.0,>=4.41.0 in /usr/local/lib/python3.12/dist-packages (from sentence-transfo
Requirement already satisfied: huggingface-hub>=0.20.0 in /usr/local/lib/python3.12/dist-packages (from sentence-transformer
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.12/dist-packages (from sentence-transformers) (2.9.0+
Requirement already satisfied: numpy in /usr/local/lib/python3.12/dist-packages (from sentence-transformers) (2.0.2)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.12/dist-packages (from sentence-transformers) (1.6.1)
Requirement already satisfied: scipy in /usr/local/lib/python3.12/dist-packages (from sentence-transformers) (1.16.3)
Requirement already satisfied: typing_extensions>=4.5.0 in /usr/local/lib/python3.12/dist-packages (from sentence-transforme
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages (from sentence-transformers) (4.67.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sentence-t
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->se
Requirement already satisfied: hf-xet<2.0.0,>=1.2.0 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0
Requirement already satisfied: httpx<1,>=0.23.0 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->se
Requirement already satisfied: packaging>=20.9 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->ser
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sentenc
Requirement already satisfied: shellingham in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sentenc
Requirement already satisfied: typer-slim in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sentence
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transform
Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transf
Requirement already satisfied: networkx>=2.5.1 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-trar
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformers)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.12/dist-packages (from transformers<6.0.0,>=4.41.
Requirement already satisfied: tokenizers<=0.23.0,>=0.22.0 in /usr/local/lib/python3.12/dist-packages (from transformers<6.0
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.12/dist-packages (from transformers<6.0.0,>=4.41
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.12/dist-packages (from scikit-learn->sentence-transfo
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.12/dist-packages (from scikit-learn->sentence-
Requirement already satisfied: anyio in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->huggingface-hub>=0.2
Requirement already satisfied: certifi in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->huggingface-hub>=0
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->huggingface-
Requirement already satisfied: idna in /usr/local/lib/python3.12/dist-packages (from httpx<1,>=0.23.0->huggingface-hub>=0.20
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.12/dist-packages (from httpcore==1.*->httpx<1,>=0.23.0->h
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.12/dist-packages (from sympy>=1.13.3->torch>=1.1
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2->torch>=1.11.0->sente
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.12/dist-packages (from typer-slim->huggingface-hub>=0.
```

```
from sentence_transformers import SentenceTransformer
import pandas as pd
import joblib

print("Loading dataset...")

# Load dataset
df = pd.read_csv("/content/Aicte Intership Dataset.csv")

# Clean data
df = df.fillna('').astype(str)
df.columns = df.columns.str.strip().str.lower()

print("Creating combined text...")

# Combine important fields
df["combined_text"] = (
    df["job title"] + " " +
    df["company name"] + " " +
    df["cities"] + " " +
    df["states"] + " " +
    df["stipend"] + " " +
    df["duration"]
)

print("Loading semantic AI model...")

# ⭐ This is the BRAIN
model = SentenceTransformer('all-MiniLM-L6-v2')

print("Converting internships into vectors (embeddings)...")

# Convert all internships → embeddings
embeddings = model.encode(
    df["combined_text"].tolist(),
    show_progress_bar=True
)

print("Saving AI files...")

# Save model + embeddings
joblib.dump(model, "semantic_model.pkl")
joblib.dump(embeddings, "internship_embeddings.pkl")
```

```
print("✅ SUCCESS — INDUSTRY MODEL READY 🚀")
```

```
Loading dataset...
Creating combined text...
Loading semantic AI model...
Loading weights: 100%  103/103 [00:00<00:00, 447.43it/s, Materializing param=pooler.dense.weight]
BertModel LOAD REPORT from: sentence-transformers/all-MiniLM-L6-v2
Key                     | Status     |  |
------------------------+------------+--+-
embeddings.position_ids | UNEXPECTED |  |

Notes:
- UNEXPECTED    :can be ignored when loading from different task/architecture; not ok if you expect identical arch.
Converting internships into vectors (embeddings)...
Batches: 100%  70/70 [00:47<00:00, 2.98it/s]
Saving AI files...
✅ SUCCESS — INDUSTRY MODEL READY 🚀
```

```
from sentence_transformers import SentenceTransformer
import numpy as np

model = joblib.load("semantic_model.pkl")
internship_embeddings = joblib.load("internship_embeddings.pkl")
```

```
from sklearn.metrics.pairwise import cosine_similarity

def semantic_recommend(resume_text, top_n=5):

    resume_embedding = model.encode([resume_text])

    similarities = cosine_similarity(
        resume_embedding,
        internship_embeddings
    )[0]

    top_indices = similarities.argsort()[-top_n:][::-1]

    results = df.iloc[top_indices].copy()

    results["Match Score (%)"] = (
        similarities[top_indices] * 100
    ).round(2)

    return results[[
        'job title',
        'company name',
        'cities',
        'states',
        'stipend',
        'duration',
        'Match Score (%)'
    ]]
```

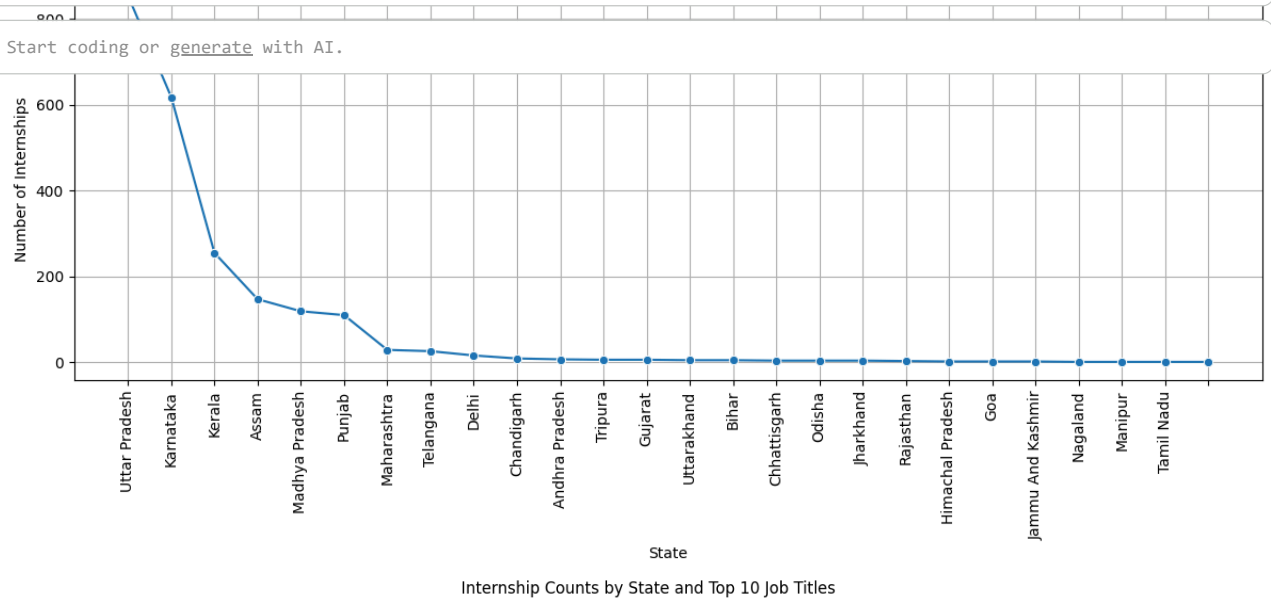
```
from sentence_transformers import SentenceTransformer
import numpy as np

model = joblib.load("semantic_model.pkl")
internship_embeddings = joblib.load("internship_embeddings.pkl")
```

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd # Ensure pandas is imported as it's used for crosstab and value_counts

# Line Plot: Number of Internships per State
state_counts = df['states'].value_counts()
plt.figure(figsize=(12, 6))
sns.lineplot(x=state_counts.index, y=state_counts.values, marker='o')
plt.title('Number of Internships per State')
plt.xlabel('State')
plt.ylabel('Number of Internships')
plt.xticks(rotation=90)
plt.grid(True)
plt.tight_layout()
plt.show()

# Heat Map: Internship Counts by State and Top 10 Job Titles
```

```
# Get the top 10 most frequent job titles
top_job_titles = df['job title'].value_counts().head(10).index.tolist()

# Filter the DataFrame to include only these top job titles
df_filtered = df[df['job title'].isin(top_job_titles)]

# Create a cross-tabulation (frequency table) for states and job titles
heatmap_data = pd.crosstab(df_filtered['states'], df_filtered['job title'])

plt.figure(figsize=(15, 10))
sns.heatmap(heatmap_data, annot=True, fmt="d", cmap="YlGnBu", linewidths=.5)
plt.title('Internship Counts by State and Top 10 Job Titles')
plt.xlabel('Job Title')
plt.ylabel('State')
plt.tight_layout()
plt.show()
```

Start coding or generate with AI.

Start coding or generate with AI.



Internship Counts by State and Top 10 Job Titles