```
# Upload the CSV file
uploaded = files.upload()

# Read the uploaded file
movies = pd.read_csv("tmdb_5000_movies.csv")

# Select and clean data
movies = movies[['title', 'overview']].dropna()

# TF-IDF vectorization
tfidf = TfidfVectorizer(stop_words='english')
tfidf_matrix = tfidf.fit_transform(movies['overview'])

# Cosine similarity matrix
cosine_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)

# Index mapping
indices = pd.Series(movies.index, index=movies['title']).drop_duplicates()

# Recommendation function
def recommend_movies(title, num_recommendations=5):
    if title not in indices:
        return f"Movie '{title}' not found."

    idx = indices[title]
```

```
    idx = indices[title]
    sim_scores = list(enumerate(cosine_sim[idx]))
    sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)[1:num_recommendations+1]
    movie_indices = [i[0] for i in sim_scores]

    return movies['title'].iloc[movie_indices]

# Test the function
movie_name = "The Dark Knight"
print(f"\nTop recommendations for '{movie_name}':\n")
print(recommend_movies(movie_name))
```

```python
import pandas as pd

# 50 Marvel movie titles
titles = [
    "Iron Man", "The Incredible Hulk", "Iron Man 2", "Thor", "Captain America: The First Avenger",
    "The Avengers", "Iron Man 3", "Thor: The Dark World", "Captain America: The Winter Soldier", "Guardians of the Galaxy",
    "Avengers: Age of Ultron", "Ant-Man", "Captain America: Civil War", "Doctor Strange", "Guardians of the Galaxy Vol. 2",
    "Spider-Man: Homecoming", "Thor: Ragnarok", "Black Panther", "Avengers: Infinity War", "Ant-Man and the Wasp",
    "Captain Marvel", "Avengers: Endgame", "Spider-Man: Far From Home", "Black Widow", "Shang-Chi and the Legend of the Ten Rings",
    "Eternals", "Spider-Man: No Way Home", "Doctor Strange in the Multiverse of Madness", "Thor: Love and Thunder", "Black Panther: Wakanda Forever",
    "Ant-Man and the Wasp: Quantumania", "Guardians of the Galaxy Vol. 3", "The Marvels", "Loki", "WandaVision",
    "Hawkeye", "Moon Knight", "Ms. Marvel", "She-Hulk", "The Falcon and the Winter Soldier",
    "What If...?", "Werewolf by Night", "Secret Invasion", "Echo", "Agatha: Darkhold Diaries",
    "Deadpool", "Deadpool 2", "Logan", "X-Men: Days of Future Past", "X-Men: Apocalypse"
]

# Dummy overviews and genres
overviews = [f"{title} is a Marvel superhero film filled with action and adventure." for title in titles]
genres = ["Action/Adventure/Superhero"] * 50

# Create and save the DataFrame
df = pd.DataFrame({
    "title": titles,
    "overview": overviews,
    "genre": genres
```

```python
# Dummy overviews and genres
overviews = [f"{title} is a Marvel superhero film filled with action and adventure." for title in titles]
genres = ["Action/Adventure/Superhero"] * 50

# Create and save the DataFrame
df = pd.DataFrame({
    "title": titles,
    "overview": overviews,
    "genre": genres
})

df.to_csv("marvel_50_movies.csv", index=False)
print("File saved as marvel_50_movies.csv")
```

```
File saved as marvel_50_movies.csv
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (1.6.1)
Requirement already satisfied: numpy>=1.23.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.15.3)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.5.0)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
```

Browse... 2 files selected.

tmdb_5000_movies.csv(application/vnd.ms-excel) - 5698602 bytes, last modified: n/a - 100% done

tmdb_5000_credits.csv(application/vnd.ms-excel) - 40044293 bytes, last modified: n/a - 100% done

```
Saving tmdb_5000_movies.csv to tmdb_5000_movies.csv
Saving tmdb_5000_credits.csv to tmdb_5000_credits.csv

Top recommendations for 'The Dark Knight':

3                         The Dark Knight Rises
428                             Batman Returns
3854    Batman: The Dark Knight Returns, Part 2
299                             Batman Forever
1359                                   Batman
Name: title, dtype: object
```