Vendor Qualification System (Peter Le)

Required Libraries
pip install pandas
pip install scikit-learn
pip install flask
pip install requests

File Overview
main.py - ingests data and scores it
tests.py - runs tests among different categories

main.py
Task 1
- Loads vendor dataset and turns it into pandas dataframe.
- Assigns a custom category column to the right of the dataframe.
    - Considers descriptions with keywords to classify different categories.
- Combines features and description into a singular column to be vectorized.

Task 2
- Computes TF-TDF vectors using TfidfVectorizer library.
    - Gives numerical weights of vendors using popular vectorization technique.
- Converts user query into query vector.
- Computes cosine similarity between query and vendors.
    - Angle between two vectors determines correlation between the vectors.
- Apply threshold filter of 0.01 (best threshold out of experimentation).
    - If no results appear, set the threshold to 0.

Task 3
- Compute a composite score.
    - Apply a 0.7 rating weight and 0.3 similarity weight (popular heuristic weights that value rating).
- Sort vendors by score (descending).
- Return top 10 vendors

Task 4
- Created flask framework.
- Endpoint - POST /vendor_qualification.

- Used port 8000 (5000 did not work with my macbook).

Testing
Two Ways to Test
1. python tests.py
   a. Sends HTTP Post requests to Flask API with requests library.
   b. Sends post request with test inputs
   c. Parses json response
   d. Returns number of results with vendor names, classified custom categories, score, and similarity
   e.

```
(base) peternle@Peters-MBP vendorqualification % python tests.py

=== Category: crm software ===
Input: {'software_category': 'crm software', 'capabilities': ['automation', 'integration']}
Number of results: 10
1. Solid Performers CRM | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 3.449 | Similarity: 0.064 (related to crm software according to description and category)
2. Fireberry | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 3.373 | Similarity: 0.042 (related to crm software according to description and category)
3. Breakcold | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 3.306 | Similarity: 0.054 (related to crm software according to description and category)
4. EspoCRM | Category: crm software | Custom Categories: crm software | Score: 3.244 | Similarity: 0.079 (related to crm software according to description and category)
5. AllClients | Category: crm software | Custom Categories: crm software | Score: 3.243 | Similarity: 0.078 (related to crm software according to description and category)
6. Zurmo | Category: crm software | Custom Categories: crm software | Score: 3.236 | Similarity: 0.053 (related to crm software according to description and category)
7. Prospect CRM | Category: crm software | Custom Categories: crm software | Score: 3.234 | Similarity: 0.046 (related to crm software according to description and category)
8. Pipeliner CRM | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 3.23 | Similarity: 0.034 (related to crm software according to description and category)
9. Efficy CRM | Category: crm software | Custom Categories: crm software, accounting & finance software, erp software, hr software | Score: 3.161 | Similarity: 0.036 (related to crm software according to description and category)
10. Freshsales | Category: crm software | Custom Categories: crm software, accounting & finance software, erp software | Score: 3.158 | Similarity: 0.027 (related to crm software according to description and category)

=== Category: accounting & finance software ===
Input: {'software_category': 'accounting & finance software', 'capabilities': ['budgeting']}
Number of results: 8
1. Solid Performers CRM | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 3.43 | Similarity: 0.0 (related to accounting & finance software according to description and category)
2. Fireberry | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 3.36 | Similarity: 0.0 (related to accounting & finance software according to description and category)
3. Breakcold | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 3.29 | Similarity: 0.0 (related to accounting & finance software according to description and category)
4. Pipeliner CRM | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 3.22 | Similarity: 0.0 (related to accounting & finance software according to description and category)
5. Efficy CRM | Category: crm software | Custom Categories: crm software, accounting & finance software, erp software, hr software | Score: 3.15 | Similarity: 0.0 (related to accounting & finance software according to description and category)
6. Freshsales | Category: crm software | Custom Categories: crm software, accounting & finance software, erp software | Score: 3.15 | Similarity: 0.0 (related to accounting & finance software according to description and category)
7. FinCRM | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 3.01 | Similarity: 0.0 (related to accounting & finance software according to description and category)
8. Kommo | Category: crm software | Custom Categories: crm software, accounting & finance software | Score: 2.59 | Similarity: 0.0 (related to accounting & finance software according to description and category)

=== Category: erp software ===
Input: {'software_category': 'erp software', 'capabilities': ['inventory']}
Number of results: 2
1. Efficy CRM | Category: crm software | Custom Categories: crm software, accounting & finance software, erp software, hr software | Score: 3.15 | Similarity: 0.0 (related to erp software according to description and category)
2. Freshsales | Category: crm software | Custom Categories: crm software, accounting & finance software, erp software | Score: 3.15 | Similarity: 0.0 (related to erp software according to description and category)
(base) peternle@Peters-MBP vendorqualification %
```

2. curl -X POST http://localhost:8000/vendor_qualification -H "Content-Type: application/json" -d '{"software_category": "Accounting & Finance Software", "capabilities": ["Budgeting"]}'
   a. 8000 instead of 5000

b.

```
(base) peternle@Peters-MBP vendorqualification % curl -X POST http://localhost:8000/vendor_qualification -H "Content-Type: application/json" -d '{"software_category": "Accounting & Finance Software", "c
apabilities": ["Budgeting"]}'
{
  "results": [
    {
      "product_name": "Solid Performers CRM",
      "main_category": "crm software",
      "custom_categories": [
        "crm software",
        "accounting & finance software"
      ],
      "score": 3.43,
      "similarity": 0.0
    },
    {
      "product_name": "Fireberry",
      "main_category": "crm software",
      "custom_categories": [
        "crm software",
        "accounting & finance software"
      ],
      "score": 3.36,
      "similarity": 0.0
    },
    {
      "product_name": "Breakcold",
      "main_category": "crm software",
      "custom_categories": [
        "crm software",
        "accounting & finance software"
      ],
      "score": 3.29,
      "similarity": 0.0
    },
    {
      "product_name": "Pipeliner CRM",
      "main_category": "crm software",
      "custom_categories": [
        "crm software",
        "accounting & finance software"
      ],
      "score": 3.2199999999999998,
      "similarity": 0.0
    },
    {
      "product_name": "Efficy CRM",
      "main_category": "crm software",
      "custom_categories": [
        "crm software",
        "accounting & finance software",
        "erp software",
        "hr software"
      ],
      "score": 3.15,
      "similarity": 0.0
    },
    {
      "product_name": "Freshsales",
      "main_category": "crm software",
      "custom_categories": [
        "crm software",
        "accounting & finance software",
        "erp software"
      ],
      "score": 3.15,
      "similarity": 0.0
    },
    {
      "product_name": "FinCRM",
      "main_category": "crm software",
      "custom_categories": [
        "crm software",
        "accounting & finance software"
      ],
      "score": 3.01,
      "similarity": 0.0
    },
    {
      "product_name": "Kommo",
      "main_category": "crm software",
      "custom_categories": [
        "crm software",
        "accounting & finance software"
      ],
      "score": 2.59,
      "similarity": 0.0
    }
  ]
}
(base) peternle@Peters-MBP vendorqualification % 
```

Challenges
- Finding a fitting threshold that would guarantee results.
- Finding fitting features and experimenting with them.
- Learning Flask API
- Vectorization research

Potential Improvements
- Experiment more with the rating and similarity weights.
- Docker support.
- More unit tests.
- Possible dynamic threshold.