https://github.com/Thocook/Arbitrage_Finder/blob/main/Arbitrage_Finder.ipynb

Arbitrage Finder

2023/12/29

A Birds Eye View

Utilizing: Google Search, Web Scraping, Vectorization, and Weighted Cosine Similarity Ranking

Developed a web scraper for Amazon and Alibaba to identify arbitrage opportunities. The scraper extracts product details, including title, price, brand, and ratings, from both platforms. Utilizing scikit learn, it vectorizes the brand and title, and employs cosine similarity to find the most similar products.

1.1 Create Link Lists

```
from googlesearch import search

query = "bathroom scale site:amazon.ca inurl:dp"

# Perform a Google search and get the URLs
amazon_links = list(search(query, num=50, stop=300, pause=2))
```

Using Google Search, take the user's query input and generate a list of Amazon and Alibaba product links. The Amazon list is larger because it produces more faulty links that need to be filtered out compared to Alibaba.

1.2 Scrape Amazon & Alibaba

```
import pandas as pd
from bs4 import BeautifulSoup
```

- Get the data from amazon and Alibaba for all of their associated links

Amazon: Title, Price, Brand, Vendor, Colour, Special Features, Ratings,Number of reviews, URL
Alibaba: Title, Price_ali, Product, Brand_Name, Store, Rating_ali, Link_ali

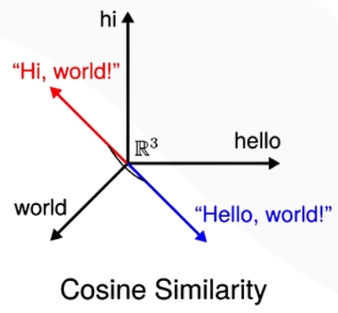
1.3 Correlation Matrix

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

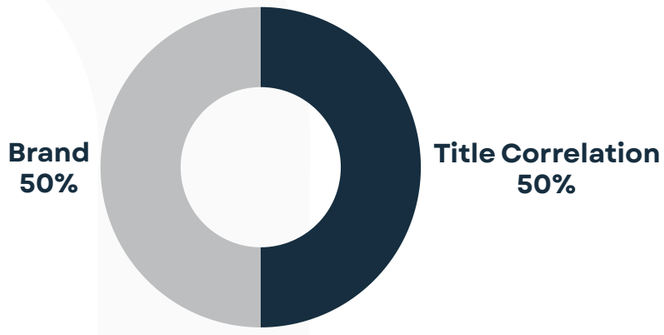
1. Transformed text data into vectorized points for brand and title.
2. Computed cosine similarity for each matrix.
3. Combined matrices with equal weighting (50% brand, 50% title).
4. Sorted and ranked scores to identify the most similar items.

Interpretation of Results:

- Brand similarity contributes 50%, title similarity contributes 50%.
- A score exceeding 80% generally signifies identical brand and closely resembling title.



Cosine Similarity



1.4 Top Result For Bathroom Scales

Correlation Score: **0.89%**
Arbitrage Opportunity: **$65 CAD**