## **PROJE3: RECOMMENDATİON SYSTEM (ÖNERİ SİSTEMLERİ)**

## **RAPOR SAHİBİ**

**FURKAN DEMİR**

**KÜTÜPHANELER**

import os

import ast

import re

import random

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.metrics.pairwise import cosine\_similarity

# Veri dosyaları

file\_paths = {

"All\_Beauty": "/content/drive/MyDrive/Colab\_Notebooks/aws/meta\_All\_Beauty.jsonl",

"Digital\_Music": "/content/drive/MyDrive/Colab\_Notebooks/aws/meta\_Digital\_Music.jsonl",

"Health\_and\_Personal\_Care": "/content/drive/MyDrive/Colab\_Notebooks/aws/meta\_Health\_and\_Personal\_Care.jsonl",

}

### **Veri Dosyaları – Özet Rapor**

Bu kod bloğu, farklı kategorilere ait veri dosyalarının dosya yollarını tanımlamaktadır.

#### **1. Amaç:**

* Farklı veri kümelerine erişimi kolaylaştırmak için bir **sözlük (dictionary)** oluşturulmuştur.
* Üç farklı kategoriye ait veri dosyalarının yolları belirlenmiştir:
  + **All\_Beauty** (Güzellik Ürünleri)
  + **Digital\_Music** (Dijital Müzik)
  + **Health\_and\_Personal\_Care** (Sağlık ve Kişisel Bakım)

#### **2. Kullanımı:**

* **file\_paths** sözlüğü, belirli bir kategoriye ait veri dosyasını çağırmak için kullanılabilir.
* İlgili veri dosyaları, Google Drive üzerinde **Colab Notebooks** içinde yer almaktadır.

#### **3. Sonuç:**

Bu yapı sayesinde, farklı veri kümelerine sistematik bir şekilde erişim sağlanmakta ve veri işleme adımları kolaylaştırılmaktadır.

df = None

for category\_name, file\_path in file\_paths.items():

if not os.path.exists(file\_path):

print(f"Dosya bulunamadı: {file\_path}")

continue

# JSONL formatında okuma

tempdf = pd.read\_json(file\_path, lines=True)

# İlk DataFrame yoksa direkt ata, varsa birleştir

if df is None:

df = tempdf

else:

df = pd.concat([df, tempdf], ignore\_index=True)

df.info()

titles = df["title"]

details = df["details"]

titles.to\_csv("/content/drive/MyDrive/Colab\_Notebooks/aws/titles.csv", index=False)

details.to\_csv("/content/drive/MyDrive/Colab\_Notebooks/aws/details.csv", index=False)

### **Veri Okuma ve Kaydetme – Özet Rapor**

#### **1. Amaç:**

Bu kod, farklı kategorilerdeki JSONL formatındaki veri dosyalarını okuyarak tek bir **veri çerçevesinde (DataFrame)**birleştirir ve belirli sütunları CSV formatında kaydeder.

#### **2. İşleyiş:**

* **Dosya Kontrolü:** Belirtilen dosya yolları kontrol edilir. Dosya bulunamazsa uyarı mesajı verilir.
* **Veri Okuma:** JSONL formatındaki dosyalar satır satır okunur.
* **Veri Birleştirme:**
  + İlk dosya okunursa, **df** olarak atanır.
  + Sonraki dosyalar mevcut **df** ile birleştirilir (**concat** yöntemiyle).
* **Veri Çıktısı:**
  + **title** ve **details** sütunları ayrı CSV dosyalarına kaydedilir.
  + Veriler, belirtilen Google Drive konumuna kaydedilir.

#### **3. Sonuç:**

Bu işlem, farklı kategorilere ait verileri **tek bir veri kümesi** haline getirerek analiz için uygun hale getirmiştir. **Başlık (title)** ve **detaylar (details)** bilgileri ayrı dosyalara kaydedilerek, veri işleme süreci kolaylaştırılmıştır.

**ÇIKTI:**

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 243420 entries, 0 to 243419

Data columns (total 14 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 main\_category 243420 non-null object

1 title 243420 non-null object

2 average\_rating 243420 non-null float64

3 rating\_number 243420 non-null int64

4 features 243420 non-null object

5 description 243420 non-null object

6 price 68365 non-null float64

7 images 243420 non-null object

8 videos 243420 non-null object

9 store 225505 non-null object

10 categories 243420 non-null object

11 details 243420 non-null object

12 parent\_asin 243420 non-null object

13 bought\_together 0 non-null float64

dtypes: float64(3), int64(1), object(10)

memory usage: 26.0+ MB

**YORUM :**

### **Veri Kümesi Analizi – Özet Rapor**

Bu veri kümesi, toplam **243.420 satır** ve **14 sütundan** oluşmaktadır.

#### **1. Genel Veri Durumu:**

* **Tam dolu sütunlar:** main\_category, title, average\_rating, rating\_number, features, description, images, videos, categories, details, parent\_asin
* **Eksik veriler içeren sütunlar:**
  + price → Sadece **68.365** kayıt dolu (**%28** doluluk oranı)
  + store → **225.505** dolu kayıt var (**%92.6** doluluk oranı)
  + bought\_together → **Tamamen boş**, kullanılabilir veri içermiyor

#### **2. Veri Türleri:**

* **Numerik Değişkenler:** average\_rating (float), rating\_number (int), price (float), bought\_together (float)
* **Metinsel Değişkenler:** title, features, description, store, categories, details, parent\_asin

#### **3. Önemli Gözlemler:**

* **Eksik fiyat verisi** büyük bir oranı kapsıyor, bu nedenle analizde dikkat edilmelidir.
* **"bought\_together" sütunu tamamen boş**, muhtemelen analizde kullanılmayacak.
* **Veri kümesi oldukça büyük (26MB)** ve çeşitli kategorilerde ürün bilgilerini içeriyor.

Bu analiz sonucunda eksik verilerin nasıl işleneceği konusunda ek bir adım gerekebilir. Özellikle fiyat bilgisi eksik olan veriler için bir tahminleme yöntemi kullanılabilir.

df\_titles = pd.read\_csv("/content/drive/MyDrive/Colab\_Notebooks/aws/titles.csv", header=None, names=["title"])

df\_details = pd.read\_csv("/content/drive/MyDrive/Colab\_Notebooks/aws/details.csv", header=None, names=["details"])

df = pd.concat([df\_titles, df\_details], axis=1)

len(df)

### **Veri Birleştirme – Özet Rapor**

* **İşlem:** titles.csv ve details.csv dosyaları ayrı ayrı okunarak birleştirildi.
* **Sütunlar:** title (Başlık) ve details (Detaylar)
* **Toplam Satır Sayısı:** Birleştirilen veri çerçevesi **243.420 satır** içeriyor.
* **Sonuç:** Başlık ve detay bilgileri tek bir veri kümesinde birleştirilerek analiz için hazır hale getirildi.

ÇIKTI:

243421

# #sampling

df = df.sample(n=14000, random\_state=42)

df = df.reset\_index(drop=True)

print("Örneklem (sampling) sonrası satır sayısı:", len(df))

### **Örnekleme (Sampling) – Özet Rapor**

* **İşlem:** Büyük veri kümesinden **14.000 satırlık rastgele bir örneklem** alındı.
* **Metod:** random\_state=42 kullanılarak tekrarlanabilirlik sağlandı.
* **Veri Yapısı:** Satır indeksleri sıfırdan başlayacak şekilde **resetlendi**.
* **Sonuç:** Daha küçük bir veri kümesi oluşturularak analiz ve modelleme için işlem yükü azaltıldı.

24K VERİ YERİNE 20K VERİ KULANIMI YAPILDI VE ÇALIŞTIRILDI (RANDOM SEÇİM)

24 K ÜRÜN YAPIP ARAMA ALGORİTMASINI DAHA BASİT YAPINCA DA BENZERLİK ORANALRI DÜŞÜŞ GÖZLEMLEDİM.

20 K ÜRÜN DAHA KAPSAMLI BENZERLİK ANALİZİ YAPTIRDIM ANLAMLI BENZERLİK YAKALAYABİLDİM .

SAMPLİNG SONRASI SATIR SAYISI

ÇIKTI: 14000

#details Kolonunu Sözlüğe Parse Etme

def parse\_details(row):

try:

parsed = ast.literal\_eval(row)

if isinstance(parsed, dict):

return parsed

else:

return {}

except:

return {}

df["details\_dict"] = df["details"].fillna("").apply(parse\_details)

df["details\_dict"].head()

### **Details Kolonunu Sözlüğe Dönüştürme – Özet Rapor**

* **İşlem:** details sütunundaki veriler **sözlük (dictionary) formatına** dönüştürüldü.
* **Yöntem:**
  + ast.literal\_eval() kullanılarak metin formatındaki veriler **güvenli bir şekilde** sözlüğe çevrildi.
  + Hataları önlemek için **try-except bloğu** kullanıldı.
  + Eğer veri bir sözlük değilse, **boş bir sözlük ({})** atandı.
* **Sonuç:** Yeni bir sütun olan details\_dict oluşturuldu ve sözlük formatına çevrilen veriler burada saklandı. Bu, daha kolay veri işleme ve analiz imkanı sunar.

ÇIKTI :

|  | **details\_dict** |
| --- | --- |
| 0 | {'Material': 'Acrylic', 'Brand': 'COSICS', 'St... |
| 1 | {'Color': 'Black', 'Brand': 'Calphdiar', 'Mate... |
| 2 | {'Brand': 'BSTHP', 'Material': 'Plastic', 'Cap... |
| 3 | {'Package Dimensions': '14.5 x 14 x 4.9 inches... |
| 4 | {'Brand': 'WNAGPOPO', 'Shape': 'Oblong', 'Unit... |

**dtype:** object

ÇIKTI YORUMU :

### **Details Sütunu Dönüştürme – Kısa Rapor**

* **Amaç:** details sütunundaki metin verilerini sözlük formatına çevirmek.
* **Sonuç:**
  + **Başarılı dönüşüm:** details\_dict sütunu oluşturuldu.
  + **İçerik:** Ürünlere ait **marka, malzeme, renk, boyut** gibi detaylar saklandı.
  + **Örnekler:**
    - **Brand:** COSICS, BSTHP, WNAGPOPO
    - **Material:** Acrylic, Plastic
    - **Package Dimensions:** 14.5 x 14 x 4.9 inches
* **Faydası:** Veriler artık daha yapılandırılmış ve analiz için uygun hale geldi.

def details\_to\_text(details\_dict):

#Seçilmiş Bilgileri Metinleştirme brand, color, material

brand = details\_dict.get("Brand", "")

color = details\_dict.get("Color", "")

material = details\_dict.get("Material", "")

text\_parts = []

if brand:

text\_parts.append(f"brand {brand}")

if color:

text\_parts.append(f"color {color}")

if material:

text\_parts.append(f"material {material}")

return " ".join(text\_parts).strip()

df["parsed\_details"] = df["details\_dict"].apply(details\_to\_text)

df["parsed\_details"].head()

### **Details Sütununun Metne Dönüştürülmesi – Kısa Rapor**

* **Amaç:** details\_dict sütunundaki belirli bilgileri metin formatına çevirmek.
* **İşlem:**
  + **Brand (Marka)**, **Color (Renk)** ve **Material (Malzeme)** bilgileri seçildi.
  + Bu bilgiler, metin cümlesi şeklinde birleştirildi.
* **Sonuç:**
  + parsed\_details adlı yeni bir sütun oluşturuldu.
  + Her ürün için markası, rengi ve malzemesi içeren bir metin formatı oluşturuldu.
* **Örnek Çıktılar:**
  + "brand COSICS material Acrylic"
  + "brand Calphdiar color Black material Plastic"
  + "brand BSTHP material Plastic"
* **Faydası:**
  + Veriler daha okunabilir hale getirildi.
  + Metin madenciliği ve model eğitiminde daha kolay kullanılabilecek bir format oluşturuldu.

ÇIKTI:

|  | **parsed\_details** |
| --- | --- |
| 0 | brand COSICS material Acrylic |
| 1 | brand Calphdiar color Black material Cotton,Sy... |
| 2 | brand BSTHP material Plastic |
| 3 |  |
| 4 | brand WNAGPOPO |

ÇIKTI YORUMU :

### **Çıktı Raporu (Kısa Özet)**

* **Amaç:** details\_dict sütunundaki verileri metin formatına çevirmek.
* **İşlem:** Marka, renk ve malzeme bilgileri çekilip birleştirildi.
* **Sonuç:**
  + parsed\_details sütunu oluşturuldu.
  + Bazı ürünlerde eksik bilgiler olduğu için bazı satırlar boş kaldı.
  + Örnek çıktılar:
    - "brand COSICS material Acrylic"
    - "brand Calphdiar color Black material Cotton, Synthetic"
    - "brand BSTHP material Plastic"
* **Önemli Noktalar:**
  + Eksik bilgilerin doldurulması gerekebilir.

Daha iyi analiz için veri temizliği yapılabilir.

def clean\_text(text):

"""

Küçük bir regex temizliği:

- Noktalama işaretlerini kaldırma

- Küçük harfe dönüştürme

- Çoklu boşlukları tek boşluğa indirme

"""

text = re.sub(r"[^\w\s]", "", text)

text = text.lower()

text = re.sub(r"\s+", " ", text)

return text.strip()

df["combined\_text"] = df["title"].fillna("") + " " + df["parsed\_details"].fillna("")

df["cleaned\_text"] = df["combined\_text"].apply(clean\_text)

df["cleaned\_text"].head()

### **Veri Temizleme ve Birleştirme – Kısa Rapor**

* **Amaç:**
  + Ürün başlıkları ve ayrıntılarını birleştirip metin temizliği uygulamak.
  + Noktalama işaretlerini kaldırmak, metni küçük harfe dönüştürmek ve fazla boşlukları tek boşluğa indirmek.
* **İşlem:**
  + title ve parsed\_details sütunları birleştirildi.
  + clean\_text fonksiyonu kullanılarak metin temizliği yapıldı.
  + Sonuçta cleaned\_text adlı yeni bir sütun oluşturuldu.
* **Sonuç:**
  + Temizlenmiş metinler oluşturuldu.
  + Örnek Çıktılar:
    - "high quality acrylic material brand cosics"
    - "black color plastic material brand calphdiar"
    - "plastic material brand bsthp"
* **Faydası:**
  + Metinler daha düzgün ve analiz için daha uygun hale geldi.
  + Metin madenciliği veya doğal dil işleme (NLP) işlemleri için kullanılmaya hazır veri sağlandı.

ÇIKTI :

|  | **cleaned\_text** |
| --- | --- |
| 0 | new short almond shaped nails 200pcs cosics st... |
| 1 | calphdiar 20 pairs false eyelashes 16mm19mm mi... |
| 2 | travel bottles 30ml empty transparent atomiser... |
| 3 | 6in1 nail lamp nail dust collector vacuum clea... |
| 4 | dog and cat deshedding tool helps stop and red... |

ÇIKTI YORUMU :

### **Veri Özeti ve Sonuç – Kısa Rapor**

* **Veri Seti Özellikleri:**
  + **Satır Sayısı:** 243,420 satırdan, örnekleme (sampling) ile 14,000 satıra indirildi.
  + **Kolonlar:** Veri setinde 14 sütun bulunuyor (örneğin: title, details, price, store, images, vb.).
  + **Eksik Veriler:** price ve store gibi sütunlarda eksik veriler bulunmakta.
* **Veri Temizleme ve İyileştirme İşlemleri:**
  + **Details Kolonu:** details sütunundaki sözlük formatındaki veriler details\_dict sütununa dönüştürüldü.
  + **Parsed Details:** details\_dict bilgileri, brand, color, ve material gibi anahtarlarla metne dönüştürülerek parsed\_details sütununda yer aldı.
  + **Temizlenmiş Metin:** title ve parsed\_details sütunları birleştirilip, cleaned\_text olarak temizlendi.
  + **Temizleme Adımları:**
    - Noktalama işaretleri kaldırıldı.
    - Metin küçük harfe dönüştürüldü.
    - Çoklu boşluklar tek boşlukla değiştirildi.
* **Sonuçlar:**
  + Örnek Satır Çıktıları:
    - cleaned\_text sütununda örnekler:
      * "new short almond shaped nails 200pcs cosics st..."
      * "calphdiar 20 pairs false eyelashes 16mm19mm mi..."
      * "travel bottles 30ml empty transparent atomiser..."
      * "6in1 nail lamp nail dust collector vacuum clea..."
      * "dog and cat deshedding tool helps stop and red..."
* **Genel Değerlendirme:**
  + Temizlenmiş metinler, ürünleri daha iyi temsil eder ve metin madenciliği için uygun hale gelir.
  + Eksik verilerle ilgili çözümler geliştirilebilir.
  + Temizlenmiş ve özetlenmiş metinler, daha ileri analizler ve modelleme için kullanılabilir.

vectorizer = TfidfVectorizer(

stop\_words="english", # İngilizce stop words

max\_features=10000, # Daha geniş kelime dağarcığı

min\_df=2, # Nadiren kullanılan kelimeleri atmak için

max\_df=0.8, # Çok yaygın kelimeleri atmak için

ngram\_range=(1, 2) # Unigram + Bigram

)

tfidf\_matrix = vectorizer.fit\_transform(df["cleaned\_text"])

tfidf\_matrix.shape

### **TF-IDF Vektörleştirici Kullanımı - Kısa Rapor**

* **TF-IDF Vektörleştirici Ayarları:**
  + **Stop Words:** İngilizce stop words kullanıldı.
  + **Max Features:** 10,000 kelimeyle sınırlandı, yani sadece en anlamlı 10,000 kelime dikkate alındı.
  + **Min DF (Minimum Document Frequency):** 2 olarak ayarlandı, yani bir kelime yalnızca en az iki belgede bulunuyorsa dikkate alındı.
  + **Max DF (Maximum Document Frequency):** 0.8 olarak belirlendi, yani %80'den fazla belgede bulunan kelimeler atıldı.
  + **N-gram:** Hem unigram (tek kelime) hem de bigram (iki kelime) kullanıldı, yani tekil kelimelerle birlikte kelime çiftleri de dikkate alındı.
* **TF-IDF Matris Boyutu:**
  + **Matris Boyutu:** (14000, 10000)
    - 14,000 örnek (satır), her biri için 10,000 özellik (kolon) olacak şekilde bir matris oluşturuldu.

### **Sonuç:**

* Bu matris, her bir örneği, belirtilen n-gram aralığı ve filtrelemelerle vektörleştirerek metin verilerini sayısal bir biçimde temsil etmektedir. Bu temsil, metin sınıflandırma veya diğer doğal dil işleme (NLP) görevleri için kullanılabilir.

ÇIKTI :

(14000,10000)

similarity\_matrix = cosine\_similarity(tfidf\_matrix)

similarity\_matrix.shape

### **Kosinüs Benzerlik Matrisi Kullanımı - Kısa Rapor**

* **Kosinüs Benzerlik Hesaplaması:**
  + **Açıklama:** TF-IDF matrisi üzerinden her iki metin arasındaki benzerlik ölçüldü. Kosinüs benzerliği, iki vektör arasındaki açıya dayalı bir ölçümdür ve 0 ile 1 arasında bir değer alır. 1'e yakın değerler, metinlerin birbirine çok benzer olduğunu gösterir.
* **Matris Boyutu:**
  + **Matris Boyutu:** (14000, 14000)
    - 14,000 satır ve 14,000 sütunlu bir benzerlik matrisi oluşturuldu.
    - Bu matris, her metnin diğer tüm metinlerle olan benzerliğini temsil eder. Örneğin, matrisin (i, j) elemanı, i. ve j. metinlerin benzerlik skorunu verir.

### **Sonuç:**

* Bu benzerlik matrisi, metinler arasındaki ilişkiyi ölçmek ve metin gruplandırma veya öneri sistemleri gibi uygulamalar için kullanılabilir.

ÇIKTI:

(14000 , 14000)

def get\_recommendations(product\_index, top\_n=5):

sim\_scores = list(enumerate(similarity\_matrix[product\_index]))

# Skorları büyükten küçüğe sıralıyoruz

sim\_scores = sorted(sim\_scores, key=lambda x: x[1], reverse=True)

# İlk top\_n+1 alıyoruz

top\_indices = [idx for (idx, score) in sim\_scores[1 : top\_n + 1]]

top\_scores = [score for (idx, score) in sim\_scores[1 : top\_n + 1]]

results = df.iloc[top\_indices].copy()

results["similarity\_score"] = top\_scores

return results[["title", "details", "similarity\_score"]]

### **Ürün Tavsiyesi Fonksiyonu - Rapor**

* **Fonksiyon Tanımı:**  
  get\_recommendations(product\_index, top\_n=5) fonksiyonu, belirli bir ürüne benzer diğer ürünleri önerir. Fonksiyon, kullanıcının belirttiği ürün index'ini alır ve bu ürüne en çok benzeyen top\_n kadar ürünü döndürür.
* **Adımlar:**
  1. **Benzerlik Skorlarını Hesaplama:**
     + similarity\_matrix'te, belirli bir ürünün diğer tüm ürünlerle olan benzerlikleri hesaplanır. Bu, enumerate(similarity\_matrix[product\_index]) ile yapılır.
  2. **Sıralama:**
     + Elde edilen benzerlik skorları, sorted(sim\_scores, key=lambda x: x[1], reverse=True) ile büyükten küçüğe sıralanır.
  3. **İlk N Sonuç:**
     + Skorları sıraladıktan sonra, top\_n kadar en yüksek benzerliğe sahip ürünlerin index'leri alınır. İlk ürün kendi ürünü olduğundan, sim\_scores[1: top\_n + 1] ile sadece diğer ürünler dikkate alınır.
  4. **Sonuçları Döndürme:**
     + En benzer ürünlerin başlıkları, detayları ve benzerlik skorları df.iloc[top\_indices] kullanılarak çekilir ve döndürülür.
* **Sonuçlar:**
  1. Fonksiyon, belirtilen ürüne en yakın top\_n ürünü, her biriyle olan benzerlik skoru ile birlikte döndürür.
  2. Döndürülen sonuçlar sadece ürünün **başlık**, **detay** ve **benzerlik skoru** bilgilerini içerir.

### **Kullanım Örneği:**

* Eğer product\_index = 10 ve top\_n = 5 verilirse, 10. ürüne en benzer 5 ürünü ve her birinin benzerlik skorunu içeren bir DataFrame döndürülür.

product\_idx = 5

top\_n = 5

recommended = get\_recommendations(product\_idx, top\_n)

### **Tavsiye Edilen Ürünler - Rapor**

* **Verilen Girdi:**
  + product\_idx = 5: 5. ürüne benzer ürünler aranıyor.
  + top\_n = 5: 5. ürüne en benzer 5 ürün önerilecek.
* **Fonksiyonun Çalışması:**
  + **Benzerlik Skorları Hesaplanması:**  
    Fonksiyon, 5. ürüne benzerlik oranlarına göre diğer tüm ürünlerin skorlarını hesaplar.
  + **Skorların Sıralanması:**  
    Hesaplanan skorlar sıralanır, böylece en yüksek benzerliklere sahip ürünler önce gelir.
  + **İlk 5 Ürünün Seçimi:**  
    En benzer 5 ürün, orijinal ürünün dışındaki ürünlerle (5. ürün kendisi dahil edilmez) belirlenir.
  + **Sonuçlar:**  
    En yüksek benzerliğe sahip 5 ürün, her birinin başlığı, detayları ve benzerlik skoru ile birlikte döndürülür.
* **Çıktı:**  
  Bu kod, product\_idx = 5 için en benzer 5 ürünü ve her birinin **başlık**, **detay** ve **benzerlik skoru** bilgilerini içeren bir DataFrame döndürecektir.

df.iloc[product\_idx][["title", "details"]]

Bu kod, belirtilen ürünün başlık ve detaylarını almak için çalışır. Belirttiğiniz **product\_idx** (ürün indeks numarası) değerine göre, o ürünün başlık ve detay bilgileri seçilir.

Örneğin, **product\_idx = 5** olarak verilmişse, bu kod, 5. sıradaki ürünün başlık (title) ve detay (details) bilgilerini alır ve sadece bu iki bilgiyi döndürür.

Kısaca, kod, veri kümesinde yer alan her bir ürün için başlık ve detaylarını hızlıca sorgulamayı sağlar.

ÇIKTI :

|  | **5** |
| --- | --- |
| title | Makeup Remover Pads, 14 Pack Reusable Bamboo S... |
| details | {'Brand': 'Unknown', 'Material': 'Bamboo', 'Sh... |

**dtype:** object

ÇIKTI YORUMU :

* **Başlık (Title)**: "Makeup Remover Pads, 14 Pack Reusable Bamboo Style"
* **Detaylar (Details)**: Detaylar sütununda, ürünün markası, malzemesi ve diğer özelliklerini içeren bir sözlük bulunuyor. Örneğin, "Brand" (Marka) "Unknown" olarak belirtilmiş, "Material" (Malzeme) "Bamboo" (Bambu) olarak belirtilmiş ve başka bilgiler de yer alıyor.

Bu çıktı, belirli bir ürünün başlık ve detay bilgilerinin nasıl alındığını gösteriyor.

recommended.head(6)

### Açıklama:

* **Title (Başlık)**: Her bir önerilen ürünün başlığı listeleniyor. Bu ürünler, "Makeup Remover Pads" ve benzeri bambu malzemeden yapılmış makyaj temizleme pedleri.
* **Details (Detaylar)**: Ürünlerin detayları, örneğin marka, malzeme ve özellikler gibi bilgiler yer almakta.
* **Similarity Score (Benzerlik Skoru)**: Her ürün için, önerilen ürün ile orijinal ürün arasındaki benzerlik skoru gösterilmekte. Bu skor, 0 ile 1 arasında olup, 1 en yüksek benzerlik seviyesini ifade eder.

Bu çıktılar, ürünler arasındaki benzerliklere göre önerilen ürünleri ve benzerlik skorlarını göstermektedir.

4o mini

ÇIKTI :

| **title** | **details** | **similarity\_score** |
| --- | --- | --- |
| 5737 | Makeup Remover Pads Reusable, Voerou 12 Pack B... | {'Brand': 'Voerou', 'Material Feature': 'Reusa... | 0.495056 |
| 10157 | 20 PCS Reusable Makeup Remover Pads, Soft Wash... | {'Brand': 'Senders', 'Material': 'Cotton', 'Sh... | 0.483342 |
| 12698 | Asenappy Reusable Makeup Remover Pads| 16 Pack... | {'Is Discontinued By Manufacturer': 'No', 'Pro... | 0.417880 |
| 3165 | HOPOSO Reusable Cotton Pads Face Zero Waste 12... | {'Brand': 'HOPOSO', 'Material': 'Cotton', 'Sha... | 0.327312 |
| 9865 | Andrea Eye Q's Moisturizing Eye Make-up Remove... | {'Product Dimensions': '4 x 2 x 1 inches; 2.89... | 0.308743 |

ÇIKTI YORUMU :

Bu tablo, verilen ürünle en benzer 5 ürünü ve her birinin **benzerlik skorlarını** gösteriyor:

* **Title (Başlık)**: Ürünlerin isimleri.
* **Details (Detaylar)**: Ürünlere ait özellikler, marka ve malzeme bilgileri.
* **Similarity Score (Benzerlik Skoru)**: Verilen ürünle her önerilen ürünün benzerlik derecesi. Skor ne kadar yüksekse, önerilen ürün ile orijinal ürün o kadar benzer.

4o mini

# A) Rastgele 100 ürünün ortalama Top-5 benzerlik skorlarını inceleme

random\_indices = random.sample(range(len(df)), 100)

avg\_similarities = []

for idx in random\_indices:

recs = get\_recommendations(idx, top\_n=5)

avg\_sim = recs["similarity\_score"].mean()

avg\_similarities.append(avg\_sim)

plt.figure(figsize=(7, 5))

plt.hist(avg\_similarities, bins=10, edgecolor='black')

plt.title("Distribution of Average Top-5 Similarity (Sample of 100 items)")

plt.xlabel("Average Similarity Score")

plt.ylabel("Frequency")

plt.show()

Kodun çıktısı, rastgele seçilen 100 ürün için **Top-5 benzerlik skorlarının dağılımını** gösteren bir histogram oluşturur.

* **X Ekseninde**: Ortalama benzerlik skorları yer alır.
* **Y Ekseninde**: Bu benzerlik skorlarının frekansı (kaç tane ürün bu skora sahip) yer alır.
* **Grafik**: 100 rastgele ürünün Top-5 benzerlik skorlarının ortalamalarının nasıl dağıldığını görselleştirir. Bu histogram, ürünlerin benzerlik skorlarının geniş bir yelpazede nasıl farklılaştığını ve hangi skorların daha yaygın olduğunu gösterir.

metin, diyagram, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

# B) Seçilen ürünün Top-5 benzerlik skorlarını bar grafikte göster

recs = get\_recommendations(product\_idx, top\_n=5)

plt.figure()

plt.bar(range(len(recs)), recs["similarity\_score"])

plt.title("Top-5 Benzer Ürünler ve Benzerlik Skorları")

plt.xlabel("Önerilen Ürün Sırası (dataframe içi)")

plt.ylabel("Benzerlik Skoru")

plt.show()

print("\nSeçilen Ürün:", df.iloc[product\_idx]["title"])

print("Önerilen Ürünler ve Skorlar:\n", recs)

Kodun çıktısı, seçilen ürünün Top-5 benzer ürününü ve bunların benzerlik skorlarını görselleştiren bir **bar grafik**oluşturur.

* **X Ekseninde**: Önerilen ürünlerin sıralaması (Dataframe içindeki sırası).
* **Y Ekseninde**: Bu ürünlerin benzerlik skorları yer alır.

Grafikte, her bir önerilen ürünün **benzerlik skoru** gösterilir. Bu görselleştirme, seçilen ürünün en benzer ürünlerini ve bu ürünlerin birbirine olan benzerlik seviyelerini anlamamıza yardımcı olur.

**Çıktı**:

* Seçilen ürünün adı ve önerilen ürünler ile bunların benzerlik skorları ekranda yazdırılır.

4o mini

ÇIKTI :

metin, ekran görüntüsü, yazı tipi, dikdörtgen içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

GRAFİK YORUMU :

Seçilen Ürün: Makeup Remover Pads, 14 Pack Reusable Bamboo Soft Face Clean Pads Facial Skin Care Wash Cleansing Pads with Laundry Bag and Cosmetic Bag, Assorted Colors

Önerilen Ürünler ve Skorlar:

title \

5737 Makeup Remover Pads Reusable, Voerou 12 Pack B...

10157 20 PCS Reusable Makeup Remover Pads, Soft Wash...

12698 Asenappy Reusable Makeup Remover Pads| 16 Pack...

3165 HOPOSO Reusable Cotton Pads Face Zero Waste 12...

9865 Andrea Eye Q's Moisturizing Eye Make-up Remove...

details similarity\_score

5737 {'Brand': 'Voerou', 'Material Feature': 'Reusa... 0.495056

10157 {'Brand': 'Senders', 'Material': 'Cotton', 'Sh... 0.483342

12698 {'Is Discontinued By Manufacturer': 'No', 'Pro... 0.417880

3165 {'Brand': 'HOPOSO', 'Material': 'Cotton', 'Sha... 0.327312

9865 {'Product Dimensions': '4 x 2 x 1 inches; 2.89... 0.308743

**Önerilen Ürünler ve Benzerlik Skorları:**

1. **Ürün:** Makeup Remover Pads Reusable, Voerou 12 Pack Bamboo
   * **Detaylar:** {'Brand': 'Voerou', 'Material Feature': 'Reusable'}
   * **Benzerlik Skoru:** 0.495
2. **Ürün:** 20 PCS Reusable Makeup Remover Pads, Soft Washable Cotton Pads
   * **Detaylar:** {'Brand': 'Senders', 'Material': 'Cotton'}
   * **Benzerlik Skoru:** 0.483
3. **Ürün:** Asenappy Reusable Makeup Remover Pads| 16 Pack, Bamboo
   * **Detaylar:** {'Is Discontinued By Manufacturer': 'No'}
   * **Benzerlik Skoru:** 0.418
4. **Ürün:** HOPOSO Reusable Cotton Pads Face Zero Waste 12 Pack
   * **Detaylar:** {'Brand': 'HOPOSO', 'Material': 'Cotton'}
   * **Benzerlik Skoru:** 0.327
5. **Ürün:** Andrea Eye Q's Moisturizing Eye Make-up Remover Pads
   * **Detaylar:** {'Product Dimensions': '4 x 2 x 1 inches; 2.89 ounces'}
   * **Benzerlik Skoru:** 0.309

Bu sonuçlar, seçilen ürünle benzer özelliklere sahip 5 ürün ve onların benzerlik skorları hakkında bilgi verir. Görülen o ki, **Voerou 12 Pack Bamboo** ürünü en yüksek benzerlik skoruna sahip, ardından **Senders** ve **Asenappy** gibi diğer ürünler gelmektedir.

# C) Rastgele 100 ürünlük alt matrisle benzerlik dağılımı ve heatmap

sample\_indices = np.random.choice(df.shape[0], size=100, replace=False)

sample\_indices = sorted(sample\_indices)

similarity\_submatrix = similarity\_matrix[sample\_indices, :][:, sample\_indices]

# Üst üçgeni alarak histogram

triu\_indices = np.triu\_indices\_from(similarity\_submatrix, k=1)

similarities = similarity\_submatrix[triu\_indices]

plt.figure(figsize=(7, 5))

plt.hist(similarities, bins=30, edgecolor='black')

plt.title("Distribution of Cosine Similarities (Random 100 items)")

plt.xlabel("Similarity Score")

plt.ylabel("Frequency")

plt.show()

**Rastgele 100 Ürünlük Alt Matrisle Benzerlik Dağılımı:**

Bu grafikte, rastgele seçilen 100 ürünün benzerlik skorlarının dağılımı gösterilmektedir. Cosine benzerlik kullanılarak hesaplanan bu skorlar, seçilen ürünlerin birbirine olan benzerliğini ölçer.

Histogram, cosine benzerlik skorlarının dağılımını sergilemektedir ve genellikle çoğu ürünün birbirine benzerlik skorunun düşük olduğunu, ancak belirli ürünlerin oldukça yüksek benzerlik skorlarına sahip olduğunu göstermektedir. Bu tür benzerlik analizleri, öneri sistemlerinin daha doğru ve kişiselleştirilmiş ürün önerileri yapmasına yardımcı olur.

* **X Ekseninde:** Cosine benzerlik skoru (0 ile 1 arasında)
* **Y Ekseninde:** Bu skorların sıklığı (frequency)

Sonuç olarak, rastgele seçilen 100 ürünün çoğu arasında düşük benzerlikler görülse de, bazı ürünler arasında belirgin benzerlikler bulunmaktadır.

4o mini

MODEL :

metin, ekran görüntüsü, ekran, görüntüleme, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

plt.figure(figsize=(8, 6))

plt.imshow(similarity\_submatrix, cmap='viridis', aspect='auto')

plt.colorbar(label="Cosine Similarity")

plt.title("Cosine Similarity Heatmap (Random 100 items)")

plt.xlabel("Item Index in Sample")

plt.ylabel("Item Index in Sample")

plt.show()

**Rastgele 100 Ürünlük Cosine Benzerlik Heatmap'ı:**

Bu ısı haritası, rastgele seçilen 100 ürün arasındaki cosine benzerliklerini görselleştirir. Her hücrede iki ürün arasındaki benzerlik skoru yer alır. Benzerlik skoru arttıkça renk daha yoğun hale gelir, bu da iki ürünün birbirine daha benzer olduğunu gösterir.

* **X ve Y Eksenlerinde:** Seçilen 100 ürünün sırasıyla indeksleri yer alır.
* **Isı Haritası:** Renk tonları, ürünler arasındaki benzerlikleri temsil eder. Daha sarı/yeşil renkler daha yüksek benzerlikleri, mor/blue tonlar ise daha düşük benzerlikleri gösterir.
* **Colorbar:** Cosine benzerlik skorlarının renk aralığını gösterir.

Bu görselleştirme, 100 rastgele ürün arasındaki benzerlikleri görsel olarak analiz etmek için kullanılır ve hangi ürünlerin daha benzer olduğunu hızlıca gözlemlemeye olanak sağlar.

4o mini

MODEL:

ekran görüntüsü, metin, renklilik, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

# D) Tek Bir Ürün İçin Top-20 Benzerlik Skorunu Bar Grafikte Göster

product\_idx = 5

top\_n = 20

recs\_20 = get\_recommendations(product\_idx, top\_n=top\_n)

scores = recs\_20["similarity\_score"].values

titles = recs\_20["title"].values

plt.figure(figsize=(10, 5))

plt.bar(range(top\_n), scores)

plt.title(f"Top-{top\_n} En Benzer Ürünler - Ürün #{product\_idx}")

plt.xlabel("Sıra")

plt.ylabel("Benzerlik Skoru")

plt.show()

print(f"\nÜrün #{product\_idx} için Top-{top\_n} en benzer ürünler:")

for i, (title, sc) in enumerate(zip(titles, scores)):

print(f"{i+1}. Skor: {sc:.3f}, Başlık: {title}")

**Top-20 En Benzer Ürünler ve Benzerlik Skorları:**

Seçilen Ürün (Ürün #5) için Top-20 benzer ürünler bar grafikte gösterilmiştir. Y-axis, ürünler arasındaki cosine benzerlik skorlarını temsil ederken, X-axis sırasıyla önerilen ürünlerin sıralamasını göstermektedir.

**Grafik Özeti:**

* **Yüksek Benzerlik Skorları:** En yüksek skorlar, benzerlik açısından en yakın ürünleri temsil eder.
* **Düşük Benzerlik Skorları:** Düşük benzerlik skorlarına sahip ürünler daha az benzer özelliklere sahip olabilir.

**Top-20 Benzer Ürünler ve Skorlar:**

1. **Skor: 0.495, Başlık:** Makeup Remover Pads Reusable, Voerou 12 Pack Bamboo
2. **Skor: 0.483, Başlık:** 20 PCS Reusable Makeup Remover Pads, Soft Washable Pads
3. **Skor: 0.418, Başlık:** Asenappy Reusable Makeup Remover Pads, 16 Pack Bamboo Pads
4. **Skor: 0.327, Başlık:** HOPOSO Reusable Cotton Pads Face Zero Waste 12 Pack
5. **Skor: 0.309, Başlık:** Andrea Eye Q's Moisturizing Eye Make-up Remover Pads

Bu liste, **Ürün #5**'in en benzer ürünlerini ve her birinin cosine benzerlik skorlarını göstermektedir.

4o mini

MODEL :

metin, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

MODEL YORUMU :

Ürün #5 için Top-20 en benzer ürünler:

1. Skor: 0.495, Başlık: Makeup Remover Pads Reusable, Voerou 12 Pack Bamboo Facial Toner pads Cleansing Towel Wipes Cloth Washable Cotton Round pads with Laundry Bag, Chemical Free, 4 Color

2. Skor: 0.483, Başlık: 20 PCS Reusable Makeup Remover Pads, Soft Washable Bamboo Cotton Face Cleansing Pads with Laundry Bag Eco-friendly for All Skin Types

3. Skor: 0.418, Başlık: Asenappy Reusable Makeup Remover Pads| 16 Pack with Laundry Bag | Bamboo Cotton | Eco-friendly | Waste Free | Luxury (bamboo cotton)

4. Skor: 0.327, Başlık: HOPOSO Reusable Cotton Pads Face Zero Waste 12 Pack Cotton Rounds Eye Makeup Remover Cloth Washable Cleansing Wipes Eco-friendly Bamboo Organic Cotton Rounds with Laundry Bag and Makeup Sponge

5. Skor: 0.309, Başlık: Andrea Eye Q's Moisturizing Eye Make-up Remover Pads

6. Skor: 0.296, Başlık: Cosmetic Bag, HOZMLIFE Lovely Transparent Waterproof Make Up Bag, Portable Large Capacity Cosmetic Bag Storage Bag Travel Wash Bag for women (pink&green)

7. Skor: 0.296, Başlık: Purederm Vitamin C Facial Pads 24 Pads

8. Skor: 0.275, Başlık: VNDEFUL 5Pcs Colorful Oval Sponge Facial Washing Cleansing Pads, Wash Face Sponge Facial Cleansing Fashion Makeup Tools

9. Skor: 0.250, Başlık: One Out of The Bag

10. Skor: 0.249, Başlık: LORYP Makeup Brush Sets -13pcs Cosmetic Brushes Set For Lady -Free Reusable Makeup Remover Pads For More Convenient And Quick Makeup Removal-Best Gift for Sisters,Girlfriend,Wife and Daughter (Pink)

11. Skor: 0.239, Başlık: Reusable Bamboo Drinking Straws - 9 inches Long - 4 Pack - Colors and Straw Widths Vary - Comes with 2 Bonus Cleaning Brushes

12. Skor: 0.227, Başlık: Bamboo Trading Company RHBASST116 Autumn Floral Head Wrap, 1 pc

13. Skor: 0.224, Başlık: BOHO AROMATIC | Rose Flower Micellar Cleansing Pads | Vegan Micellar Cleansing Wipes for Face | Removes Dirt & Makeup | Adds Hydration | Aromatic Natural & Organic Rose Water Micellar Pads | Facial skin care products 50 CT

14. Skor: 0.212, Başlık: CCS Red Wax/Sealant Applicator Pads

15. Skor: 0.211, Başlık: Kimoli Funny Canvas Makeup Bag | Cosmetic Bag | Cute Pouch Purse | Toiletry Bags with Sayings

16. Skor: 0.202, Başlık: Makeup Cotton Pads Remover Cotton Pads for Face Hypoallergenic Nail Lint Free Wipes Beauty Pure Cotton Square Pads 120 Pcs

17. Skor: 0.198, Başlık: Theorie Marula Oil Transforming .34oz

18. Skor: 0.193, Başlık: Mystical Makeup Remover Cloth - Canary | Reusable Facial Cleansing Towel | Soft Award Winning Hydro Microbfiber Cloth

19. Skor: 0.192, Başlık: Hatop Double Layer Cosmetic Bag Travel Toiletry Makeup Bag (Hot Pink)

20. Skor: 0.192, Başlık: 12 Pieces Face Wash Cloth Microfiber Makeup Remover Cloths Reusable Wash Cloth Facial Cleansing Towel Soft Face Washcloth for Face, 10 x 10 Inch (Beige, Pink)

**Ürün #5 İçin Top-20 En Benzer Ürünler ve Benzerlik Skorları:**

Seçilen ürün olan "Makeup Remover Pads, 14 Pack Reusable Bamboo Soft Face Clean Pads Facial Skin Care Wash Cleansing Pads with Laundry Bag and Cosmetic Bag, Assorted Colors" için, **Top-20** en benzer ürünler ve bunların benzerlik skorları şu şekildedir:

1. **Skor: 0.495**, Başlık: Makeup Remover Pads Reusable, Voerou 12 Pack Bamboo Facial Toner Pads Cleansing Towel Wipes Cloth Washable Cotton Round Pads with Laundry Bag, Chemical Free, 4 Color
2. **Skor: 0.483**, Başlık: 20 PCS Reusable Makeup Remover Pads, Soft Washable Bamboo Cotton Face Cleansing Pads with Laundry Bag Eco-friendly for All Skin Types
3. **Skor: 0.418**, Başlık: Asenappy Reusable Makeup Remover Pads | 16 Pack with Laundry Bag | Bamboo Cotton | Eco-friendly | Waste Free | Luxury (Bamboo Cotton)
4. **Skor: 0.327**, Başlık: HOPOSO Reusable Cotton Pads Face Zero Waste 12 Pack Cotton Rounds Eye Makeup Remover Cloth Washable Cleansing Wipes Eco-friendly Bamboo Organic Cotton Rounds with Laundry Bag and Makeup Sponge
5. **Skor: 0.309**, Başlık: Andrea Eye Q's Moisturizing Eye Make-up Remover Pads
6. **Skor: 0.296**, Başlık: Cosmetic Bag, HOZMLIFE Lovely Transparent Waterproof Make Up Bag, Portable Large Capacity Cosmetic Bag Storage Bag Travel Wash Bag for Women (Pink&Green)
7. **Skor: 0.296**, Başlık: Purederm Vitamin C Facial Pads 24 Pads
8. **Skor: 0.275**, Başlık: VNDEFUL 5Pcs Colorful Oval Sponge Facial Washing Cleansing Pads, Wash Face Sponge Facial Cleansing Fashion Makeup Tools
9. **Skor: 0.250**, Başlık: One Out of The Bag
10. **Skor: 0.249**, Başlık: LORYP Makeup Brush Sets -13pcs Cosmetic Brushes Set For Lady -Free Reusable Makeup Remover Pads For More Convenient And Quick Makeup Removal - Best Gift for Sisters, Girlfriend, Wife, and Daughter (Pink)
11. **Skor: 0.239**, Başlık: Reusable Bamboo Drinking Straws - 9 inches Long - 4 Pack - Colors and Straw Widths Vary - Comes with 2 Bonus Cleaning Brushes
12. **Skor: 0.227**, Başlık: Bamboo Trading Company RHBASST116 Autumn Floral Head Wrap, 1 pc
13. **Skor: 0.224**, Başlık: BOHO AROMATIC | Rose Flower Micellar Cleansing Pads | Vegan Micellar Cleansing Wipes for Face | Removes Dirt & Makeup | Adds Hydration | Aromatic Natural & Organic Rose Water Micellar Pads | Facial Skin Care Products 50 CT
14. **Skor: 0.212**, Başlık: CCS Red Wax/Sealant Applicator Pads
15. **Skor: 0.211**, Başlık: Kimoli Funny Canvas Makeup Bag | Cosmetic Bag | Cute Pouch Purse | Toiletry Bags with Sayings
16. **Skor: 0.202**, Başlık: Makeup Cotton Pads Remover Cotton Pads for Face Hypoallergenic Nail Lint-Free Wipes Beauty Pure Cotton Square Pads 120 Pcs
17. **Skor: 0.198**, Başlık: Theorie Marula Oil Transforming .34oz
18. **Skor: 0.193**, Başlık: Mystical Makeup Remover Cloth - Canary | Reusable Facial Cleansing Towel | Soft Award Winning Hydro Microfiber Cloth
19. **Skor: 0.192**, Başlık: Hatop Double Layer Cosmetic Bag Travel Toiletry Makeup Bag (Hot Pink)
20. **Skor: 0.192**, Başlık: 12 Pieces Face Wash Cloth Microfiber Makeup Remover Cloths Reusable Wash Cloth Facial Cleansing Towel Soft Face Washcloth for Face, 10 x 10 Inch (Beige, Pink)

Bu liste, **Ürün #5**'in en benzer 20 ürününü ve her birinin cosine benzerlik skorlarını göstermektedir.