

```
# Caminhos principais
IMAGE_DIR = "/content/imagens"
CACHE_FILE = "/content/catalogo_cache.npz"
os.makedirs(IMAGE_DIR, exist_ok=True)
```

```
Requirement already satisfied: parso<0.9.0,>=0.8.4 in /usr/local/lib/python3.12/dist-packages (from jedi<0.10.0,>=python3.4.0->jupyterlab)
Requirement already satisfied: entrypoints in /usr/local/lib/python3.12/dist-packages (from jupyter-client<6.1.12->ipykernel<4.5.0,>=4.1.0,!=4.0.0,!=4.0.1,!=4.0.2,!=4.0.3,!=4.0.4,!=4.0.5,!=4.0.6,!=4.0.7,!=4.0.8,!=4.0.9,!=4.0.10,!=4.0.11,!=4.0.12,!=4.0.13,!=4.0.14,!=4.0.15,!=4.0.16,!=4.0.17,!=4.0.18,!=4.0.19,!=4.0.20,!=4.0.21,!=4.0.22,!=4.0.23,!=4.0.24,!=4.0.25,!=4.0.26,!=4.0.27,!=4.0.28,!=4.0.29,!=4.0.30,!=4.0.31,!=4.0.32,!=4.0.33,!=4.0.34,!=4.0.35,!=4.0.36,!=4.0.37,!=4.0.38,!=4.0.39,!=4.0.40,!=4.0.41,!=4.0.42,!=4.0.43,!=4.0.44,!=4.0.45,!=4.0.46,!=4.0.47,!=4.0.48,!=4.0.49,!=4.0.50,!=4.0.51,!=4.0.52,!=4.0.53,!=4.0.54,!=4.0.55,!=4.0.56,!=4.0.57,!=4.0.58,!=4.0.59,!=4.0.60,!=4.0.61,!=4.0.62,!=4.0.63,!=4.0.64,!=4.0.65,!=4.0.66,!=4.0.67,!=4.0.68,!=4.0.69,!=4.0.70,!=4.0.71,!=4.0.72,!=4.0.73,!=4.0.74,!=4.0.75,!=4.0.76,!=4.0.77,!=4.0.78,!=4.0.79,!=4.0.80,!=4.0.81,!=4.0.82,!=4.0.83,!=4.0.84,!=4.0.85,!=4.0.86,!=4.0.87,!=4.0.88,!=4.0.89,!=4.0.90,!=4.0.91,!=4.0.92,!=4.0.93,!=4.0.94,!=4.0.95,!=4.0.96,!=4.0.97,!=4.0.98,!=4.0.99,!=4.0.100,!=4.0.101,!=4.0.102,!=4.0.103,!=4.0.104,!=4.0.105,!=4.0.106,!=4.0.107,!=4.0.108,!=4.0.109,!=4.0.110,!=4.0.111,!=4.0.112,!=4.0.113,!=4.0.114,!=4.0.115,!=4.0.116,!=4.0.117,!=4.0.118,!=4.0.119,!=4.0.120,!=4.0.121,!=4.0.122,!=4.0.123,!=4.0.124,!=4.0.125,!=4.0.126,!=4.0.127,!=4.0.128,!=4.0.129,!=4.0.130,!=4.0.131,!=4.0.132,!=4.0.133,!=4.0.134,!=4.0.135,!=4.0.136,!=4.0.137,!=4.0.138,!=4.0.139,!=4.0.140,!=4.0.141,!=4.0.142,!=4.0.143,!=4.0.144,!=4.0.145,!=4.0.146,!=4.0.147,!=4.0.148,!=4.0.149,!=4.0.150,!=4.0.151,!=4.0.152,!=4.0.153,!=4.0.154,!=4.0.155,!=4.0.156,!=4.0.157,!=4.0.158,!=4.0.159,!=4.0.160,!=4.0.161,!=4.0.162,!=4.0.163,!=4.0.164,!=4.0.165,!=4.0.166,!=4.0.167,!=4.0.168,!=4.0.169,!=4.0.170,!=4.0.171,!=4.0.172,!=4.0.173,!=4.0.174,!=4.0.175,!=4.0.176,!=4.0.177,!=4.0.178,!=4.0.179,!=4.0.180,!=4.0.181,!=4.0.182,!=4.0.183,!=4.0.184,!=4.0.185,!=4.0.186,!=4.0.187,!=4.0.188,!=4.0.189,!=4.0.190,!=4.0.191,!=4.0.192,!=4.0.193,!=4.0.194,!=4.0.195,!=4.0.196,!=4.0.197,!=4.0.198,!=4.0.199,!=4.0.200,!=4.0.201,!=4.0.202,!=4.0.203,!=4.0.204,!=4.0.205,!=4.0.206,!=4.0.207,!=4.0.208,!=4.0.209,!=4.0.210,!=4.0.211,!=4.0.212,!=4.0.213,!=4.0.214,!=4.0.215,!=4.0.216,!=4.0.217,!=4.0.218,!=4.0.219,!=4.0.220,!=4.0.221,!=4.0.222,!=4.0.223,!=4.0.224,!=4.0.225,!=4.0.226,!=4.0.227,!=4.0.228,!=4.0.229,!=4.0.230,!=4.0.231,!=4.0.232,!=4.0.233,!=4.0.234,!=4.0.235,!=4.0.236,!=4.0.237,!=4.0.238,!=4.0.239,!=4.0.240,!=4.0.241,!=4.0.242,!=4.0.243,!=4.0.244,!=4.0.245,!=4.0.246,!=4.0.247,!=4.0.248,!=4.0.249,!=4.0.250,!=4.0.251,!=4.0.252,!=4.0.253,!=4.0.254,!=4.0.255,!=4.0.256,!=4.0.257,!=4.0.258,!=4.0.259,!=4.0.260,!=4.0.261,!=4.0.262,!=4.0.263,!=4.0.264,!=4.0.265,!=4.0.266,!=4.0.267,!=4.0.268,!=4.0.269,!=4.0.270,!=4.0.271,!=4.0.272,!=4.0.273,!=4.0.274,!=4.0.275,!=4.0.276,!=4.0.277,!=4.0.278,!=4.0.279,!=4.0.280,!=4.0.281,!=4.0.282,!=4.0.283,!=4.0.284,!=4.0.285,!=4.0.286,!=4.0.287,!=4.0.288,!=4.0.289,!=4.0.290,!=4.0.291,!=4.0.292,!=4.0.293,!=4.0.294,!=4.0.295,!=4.0.296,!=4.0.297,!=4.0.298,!=4.0.299,!=4.0.300,!=4.0.301,!=4.0.302,!=4.0.303,!=4.0.304,!=4.0.305,!=4.0.306,!=4.0.307,!=4.0.308,!=4.0.309,!=4.0.310,!=4.0.311,!=4.0.312,!=4.0.313,!=4.0.314,!=4.0.315,!=4.0.316,!=4.0.317,!=4.0.318,!=4.0.319,!=4.0.320,!=4.0.321,!=4.0.322,!=4.0.323,!=4.0.324,!=4.0.325,!=4.0.326,!=4.0.327,!=4.0.328,!=4.0.329,!=4.0.330,!=4.0.331,!=4.0.332,!=4.0.333,!=4.0.334,!=4.0.335,!=4.0.336,!=4.0.337,!=4.0.338,!=4.0.339,!=4.0.340,!=4.0.341,!=4.0.342,!=4.0.343,!=4.0.344,!=4.0.345,!=4.0.346,!=4.0.347,!=4.0.348,!=4.0.349,!=4.0.350,!=4.0.351,!=4.0.352,!=4.0.353,!=4.0.354,!=4.0.355,!=4.0.356,!=4.0.357,!=4.0.358,!=4.0.359,!=4.0.360,!=4.0.361,!=4.0.362,!=4.0.363,!=4.0.364,!=4.0.365,!=4.0.366,!=4.0.367,!=4.0.368,!=4.0.369,!=4.0.370,!=4.0.371,!=4.0.372,!=4.0.373,!=4.0.374,!=4.0.375,!=4.0.376,!=4.0.377,!=4.0.378,!=4.0.379,!=4.0.380,!=4.0.381,!=4.0.382,!=4.0.383,!=4.0.384,!=4.0.385,!=4.0.386,!=4.0.387,!=4.0.388,!=4.0.389,!=4.0.390,!=4.0.391,!=4.0.392,!=4.0.393,!=4.0.394,!=4.0.395,!=4.0.396,!=4.0.397,!=4.0.398,!=4.0.399,!=4.0.400,!=4.0.401,!=4.0.402,!=4.0.403,!=4.0.404,!=4.0.405,!=4.0.406,!=4.0.407,!=4.0.408,!=4.0.409,!=4.0.410,!=4.0.411,!=4.0.412,!=4.0.413,!=4.0.414,!=4.0.415,!=4.0.416,!=4.0.417,!=4.0.418,!=4.0.419,!=4.0.420,!=4.0.421,!=4.0.422,!=4.0.423,!=4.0.424,!=4.0.425,!=4.0.426,!=4.0.427,!=4.0.428,!=4.0.429,!=4.0.430,!=4.0.431,!=4.0.432,!=4.0.433,!=4.0.434,!=4.0.435,!=4.0.436,!=4.0.437,!=4.0.438,!=4.0.439,!=4.0.440,!=4.0.441,!=4.0.442,!=4.0.443,!=4.0.444,!=4.0.445,!=4.0.446,!=4.0.447,!=4.0.448,!=4.0.449,!=4.0.450,!=4.0.451,!=4.0.452,!=4.0.453,!=4.0.454
```

 $\frac{1}{8}$

```
# from google.colab import files
# uploaded = files.upload() # Upload manual do CSV (desativado)
```

```
# =====
# 📁 UPLOAD DE IMAGENS
# =====
uploaded = files.upload()
for fname in uploaded.keys():
    with open(os.path.join(IMAGE_DIR, fname), 'wb') as f:
        f.write(uploaded[fname])
print(f"📁 {len(uploaded)} imagens enviadas.")
```

📁 Escolher arquivos 25 arquivos

- **Sapato7.png**(image/png) - 100472 bytes, last modified: 05/09/2025 - 100% done
- **Sapato3.jpg**(image/jpeg) - 94575 bytes, last modified: 05/09/2025 - 100% done
- **Sapato2.jpg**(image/jpeg) - 53136 bytes, last modified: 05/09/2025 - 100% done
- **Sapato1.png**(image/png) - 141983 bytes, last modified: 05/09/2025 - 100% done
- **Sapato.png**(image/png) - 55760 bytes, last modified: 05/09/2025 - 100% done
- **Sapato.jpg**(image/jpeg) - 73632 bytes, last modified: 05/09/2025 - 100% done
- **Relogio6.png**(image/png) - 44740 bytes, last modified: 05/09/2025 - 100% done
- **Relogio5.jpg**(image/jpeg) - 105084 bytes, last modified: 05/09/2025 - 100% done
- **Relogio3.jpeg**(image/jpeg) - 92427 bytes, last modified: 05/09/2025 - 100% done
- **Relogio2.jpeg**(image/jpeg) - 185584 bytes, last modified: 05/09/2025 - 100% done
- **Relogio1.png**(image/png) - 283992 bytes, last modified: 05/09/2025 - 100% done
- **Relogio.jpg**(image/jpeg) - 48289 bytes, last modified: 05/09/2025 - 100% done
- **Perfume8.jpg**(image/jpeg) - 68794 bytes, last modified: 05/09/2025 - 100% done
- **Perfume5.png**(image/png) - 111418 bytes, last modified: 05/09/2025 - 100% done
- **Perfume3.png**(image/png) - 8550 bytes, last modified: 05/09/2025 - 100% done
- **Perfume2.png**(image/png) - 27104 bytes, last modified: 05/09/2025 - 100% done
- **Perfume1.png**(image/png) - 18192 bytes, last modified: 05/09/2025 - 100% done
- **Perfume.jpg**(image/jpeg) - 137685 bytes, last modified: 05/09/2025 - 100% done
- **Jaqueta3.jpg**(image/jpeg) - 86937 bytes, last modified: 05/09/2025 - 100% done
- **Jaqueta2.jpg**(image/jpeg) - 20020 bytes, last modified: 05/09/2025 - 100% done
- **Jaqueta1.jpg**(image/jpeg) - 24433 bytes, last modified: 05/09/2025 - 100% done
- **Jaqueta.jpg**(image/jpeg) - 39000 bytes, last modified: 05/09/2025 - 100% done
- **Blazer2.jpg**(image/jpeg) - 67639 bytes, last modified: 05/09/2025 - 100% done
- **Blazer1.jpeg**(image/jpeg) - 87064 bytes, last modified: 05/09/2025 - 100% done
- **Blazer.jpg**(image/jpeg) - 26918 bytes, last modified: 05/09/2025 - 100% done

Saving Sapato7.png to Sapato7.png
 Saving Sapato3.jpg to Sapato3.jpg
 Saving Sapato2.jpg to Sapato2.jpg
 Saving Sapato1.png to Sapato1.png
 Saving Sapato.png to Sapato.png
 Saving Sapato.jpg to Sapato.jpg
 Saving Relogio6.png to Relogio6.png
 Saving Relogio5.jpg to Relogio5.jpg
 Saving Relogio3.jpeg to Relogio3.jpeg
 Saving Relogio2.jpeg to Relogio2.jpeg
 Saving Relogio1.png to Relogio1.png
 Saving Relogio.jpg to Relogio.jpg
 Saving Perfume8.jpg to Perfume8.jpg
 Saving Perfume5.png to Perfume5.png
 Saving Perfume3.png to Perfume3.png
 Saving Perfume2.png to Perfume2.png
 Saving Perfume1.png to Perfume1.png
 Saving Perfume.jpg to Perfume.jpg
 Saving Jaqueta3.jpg to Jaqueta3.jpg
 Saving Jaqueta2.jpg to Jaqueta2.jpg
 Saving Jaqueta1.jpg to Jaqueta1.jpg
 Saving Jaqueta.jpg to Jaqueta.jpg
 Saving Blazer2.jpg to Blazer2.jpg
 Saving Blazer1.jpeg to Blazer1.jpeg
 Saving Blazer.jpg to Blazer.jpg
 📁 25 imagens enviadas.

```
# =====
# ✂️ FUNÇÕES AUXILIARES
# =====
def load_and_preprocess_image(path, size=(224, 224)):
    img = PILImage.open(path).convert('RGB')
    img = img.resize(size)
    return np.array(img) / 255.0

# Carrega modelo de extração de features
model_url = "https://tfhub.dev/google/bit/m-r50x1/1"
encoder = hub.load(model_url)

def extract_features(img_array):
    img_tensor = tf.convert_to_tensor([img_array], dtype=tf.float32)
    features = encoder(img_tensor)
    return np.array(features)[0]
```

```

# =====
# 🗄️ CACHE AUTOMÁTICO
# =====
if os.path.exists(CACHE_FILE):
    cache = np.load(CACHE_FILE, allow_pickle=True)
    vectors = list(cache['vectors'])
    names = list(cache['names'])
    print(f"📦 Cache carregado com {len(names)} itens.")
else:
    vectors, names = [], []
    print("📁 Nenhum cache encontrado, criando do zero.")

all_images = sorted(os.listdir(IMAGE_DIR))
new_images = [img for img in all_images if img not in names]

if new_images:
    print(f"➕ {len(new_images)} imagens novas encontradas.")
    for img_name in tqdm(new_images):
        img_array = load_and_preprocess_image(os.path.join(IMAGE_DIR, img_name))
        vec = extract_features(img_array)
        vectors.append(vec)
        names.append(img_name)
    np.savez(CACHE_FILE, vectors=np.array(vectors), names=np.array(names))
    print("🗄️ Cache atualizado.")
else:
    print("✅ Nenhuma imagem nova.")

df = pd.DataFrame({"name": names})
images = [load_and_preprocess_image(os.path.join(IMAGE_DIR, n)) for n in df['name']]
vectors = np.array(vectors)

🔄 📁 Nenhum cache encontrado, criando do zero.
➕ 25 imagens novas encontradas.
100%|██████████| 25/25 [00:13<00:00, 1.83it/s]
🗄️ Cache atualizado.

# =====
# 📜 HISTÓRICO DE EXCLUSÕES
# =====
deleted_items = []

def restore_item(name):
    global vectors, df, images, deleted_items
    item = next((x for x in deleted_items if x['name'] == name), None)
    if item:
        idx = min(item['idx'], len(df))
        df = pd.concat([df.iloc[:idx], pd.DataFrame({"name": [item['name']]})], df.iloc[idx:]).reset_index(drop=True)
        images.insert(idx, item['image'])
        vectors = np.insert(vectors, idx, [item['vector']], axis=0)
        PILImage.fromarray((item['image'] * 255).astype(np.uint8)).save(os.path.join(IMAGE_DIR, item['name']))
        np.savez(CACHE_FILE, vectors=vectors, names=df['name'].values)
        deleted_items = [x for x in deleted_items if x['name'] != name]
        show_recommendations(idx)

def delete_item(idx):
    global vectors, df, images, deleted_items
    del_name = df.loc[idx, 'name']
    deleted_items.append({'idx': idx, 'name': del_name, 'image': images[idx], 'vector': vectors[idx]})
    df.drop(idx, inplace=True)
    df.reset_index(drop=True, inplace=True)
    images.pop(idx)
    vectors = np.delete(vectors, idx, axis=0)
    img_path = os.path.join(IMAGE_DIR, del_name)
    if os.path.exists(img_path):
        os.remove(img_path)
    np.savez(CACHE_FILE, vectors=vectors, names=df['name'].values)
    show_restore_options()

def show_restore_options():
    if deleted_items:
        undo_btn = widgets.Button(description="🔄 Desfazer último", button_style='info')
        undo_btn.on_click(lambda b: restore_item(deleted_items[-1]['name']))
        dropdown = widgets.Dropdown(options=[x['name'] for x in deleted_items], description='Restaurar:')
        restore_btn = widgets.Button(description="Restaurar selecionado", button_style='success')
        restore_btn.on_click(lambda b: restore_item(dropdown.value))
        display(widgets.HBox([undo_btn, dropdown, restore_btn]))

# =====
# 🖼️ VITRINE INTERATIVA
# =====
def image_card_with_delete(img_array, label, idx, on_click_callback):

```

```

buf = io.BytesIO()
PILImage.fromarray((img_array * 255).astype(np.uint8)).save(buf, format='PNG')
buf.seek(0)
img_widget = widgets.Image(value=buf.getvalue(), format='png', width=150, height=150)
btn_ref = widgets.Button(description=label, layout=widgets.Layout(width='150px'))
btn_ref.on_click(lambda b: on_click_callback(idx))
btn_del = widgets.Button(description="🗑", layout=widgets.Layout(width='40px'))
btn_del.on_click(lambda b: confirm_delete(idx))
return widgets.VBox([img_widget, btn_ref, btn_del])

def confirm_delete(idx):
    del_name = df.loc[idx, 'name']
    confirm_label = widgets.Label(f"Tem certeza que deseja excluir '{del_name}'?")
    btn_yes = widgets.Button(description="Sim", button_style='danger')
    btn_no = widgets.Button(description="Não", button_style='success')
    btn_yes.on_click(lambda b: (clear_output(wait=True), delete_item(idx)))
    btn_no.on_click(lambda b: (clear_output(wait=True), show_recommendations(idx)))
    display(widgets.VBox([confirm_label, widgets.HBox([btn_yes, btn_no])]))

def show_recommendations(ref_index, top_k=5):
    clear_output(wait=True)
    if len(df) == 0:
        print("📁 Catálogo vazio.")
        show_restore_options()
        return
    ref_vector = vectors[ref_index]
    similarities = cosine_similarity([ref_vector], vectors)[0]
    top_indices = np.argsort(similarities)[-1][1:top_k+1]
    print(f"Imagem de referência: {df['name'][ref_index]}")
    display(image_card_with_delete(images[ref_index], f"Referência\n(100%)", ref_index, show_recommendations))
    print("\nMais parecidas:")
    buttons = [image_card_with_delete(images[idx], f"{df['name'][idx]}\n({similarities[idx]*100:.1f}%)", idx, show_recommendations) for
    display(widgets.HBox(buttons))
    display(search_box)
    show_restore_options()

def on_search_change(change):
    query = change['new'].strip().lower()
    matches = df[df['name'].str.lower().str.contains(query)]
    if not matches.empty:
        show_recommendations(matches.index[0], top_k=5)

search_box = widgets.Text(placeholder='Digite parte do nome...', description='Buscar:', continuous_update=False)
search_box.observe(on_search_change, names='value')

# =====
# 🚀 INICIAR SISTEMA
# =====
show_recommendations(0, top_k=5)

```

➡ Imagem de referência: Blazer.jpg



Referência (100%)



Mais parecidas:



Blazer1.jpeg (89.6%)



Blazer2.jpg (86.5%)



Jaqueta.jpg (69.4%)



Jaqueta1.jpg (64.1%)



Jaqueta3.jpg (61.8%)



Buscar:

```

# =====
# 🗖️ BUSCA POR CATEGORIA (MESMA CATEGORIA)
# =====
def show_recommendations_filtered(ref_index, candidate_indices, top_k=5):
    """
    Mostra 1 referência e as 5 mais parecidas, calculadas SOMENTE dentro
    do subconjunto candidate_indices (mesma categoria/termo).
    """
    clear_output(wait=True)
    if len(candidate_indices) == 0:
        print("Nenhum candidato encontrado para a categoria.")
        display(search_box)
        show_restore_options()
        return

    # Vetor de referência
    ref_vector = vectors[ref_index]
    # Vetores apenas dos candidatos
    cand_vectors = vectors[candidate_indices]
    sims = cosine_similarity([ref_vector], cand_vectors)[0]

    # Ordena por similaridade dentro do subconjunto e remove o próprio ref_index
    order = np.argsort(sims)[::-1]
    ordered_indices = [candidate_indices[i] for i in order if candidate_indices[i] != ref_index]
    top_indices = ordered_indices[:top_k]

    print(f"Imagem de referência: {df['name'][ref_index]}")
    display(image_card_with_delete(images[ref_index], f"Referência\n(100%)", ref_index, lambda idx: show_recommendations_filtered(idx, candidate_indices, top_k=5)))

    print("\nMais 5 parecidas (mesma categoria):")
    buttons = []
    for idx in top_indices:
        sim_percent = cosine_similarity([ref_vector], [vectors[idx]])[0][0] * 100
        buttons.append(image_card_with_delete(images[idx], f"{df['name'][idx]}\n({sim_percent:.1f}%)", idx, lambda i: show_recommendations_filtered(i, candidate_indices, top_k=5)))

    if buttons:
        display(widgets.HBox(buttons))
    else:
        print("Sem similares suficientes nessa categoria.")

    display(search_box)
    show_restore_options()

def on_search_change_same_category(change):
    """
    Busca por termo (ex.: 'relógio', 'sapato', 'blazer', 'perfume', 'jaqueta')
    e mostra 1 referência + 5 similares dentro do mesmo grupo encontrado.
    """
    query = change['new'].strip().lower()
    if not query:
        return
    matches = df[df['name'].str.lower().str.contains(query)]
    if matches.empty:
        clear_output(wait=True)
        print(f"Nenhum resultado encontrado para: {query}")
        display(search_box)
        show_restore_options()
        return

    # Índice da referência (primeira ocorrência) e subconjunto candidato
    ref_index = matches.index[0]
    candidate_indices = matches.index.tolist()
    show_recommendations_filtered(ref_index, candidate_indices, top_k=5)

# Substituí o handler antigo pelo novo (mesma categoria)
try:
    search_box.unobserve_all('value')
except Exception:
    pass
search_box.observe(on_search_change_same_category, names='value')

print("Busca atualizada: referência + 5 similares dentro da MESMA categoria digitada.")

🔄 Busca atualizada: referência + 5 similares dentro da MESMA categoria digitada.

# =====
# 🗖️ UTILIDADES
# =====

def salvar_cache():
    np.savez(CACHE_FILE, vectors=vectors, names=df['name'].values)
    print(f"Cache salvo: {CACHE_FILE}")

```

```

def rebuild_from_folder():
    """
    Recalcula TUDO a partir da pasta de imagens (útil se o cache corromper ou se trocar o encoder).
    """
    global vectors, names, df, images
    file_list = sorted(os.listdir(IMAGE_DIR))
    new_vectors, new_names = [], []
    for img_name in tqdm(file_list):
        img_array = load_and_preprocess_image(os.path.join(IMAGE_DIR, img_name))
        vec = extract_features(img_array)
        new_vectors.append(vec)
        new_names.append(img_name)

    vectors = np.array(new_vectors)
    names = new_names
    df = pd.DataFrame({"name": names})
    images = [load_and_preprocess_image(os.path.join(IMAGE_DIR, n)) for n in df['name']]
    salvar_cache()
    print("Reconstrução concluída.")

print("Utilidades carregadas: salvar_cache(), rebuild_from_folder()")

↩ Utilidades carregadas: salvar_cache(), rebuild_from_folder()

# =====
# 📁 BLOCO 12 - MATRIZ DE SIMILARIDADE ENTRE CATEGORIAS
# =====

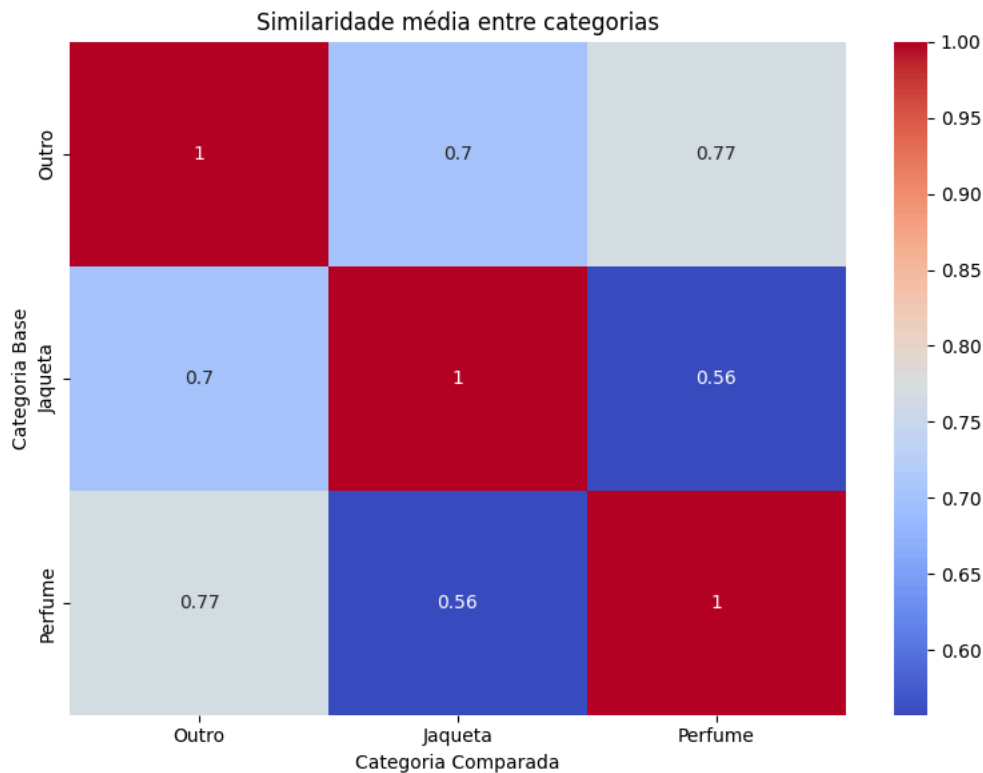
# 🐞 Atribuição automática de categorias com base no nome do arquivo
df['categoria'] = df['name'].apply(lambda x:
    'Tênis' if 'tenis' in x.lower() else
    'Perfume' if 'perfume' in x.lower() else
    'Jaqueta' if 'jaqueta' in x.lower() else
    'Outro'
)

# 📁 Geração da matriz de similaridade entre categorias
categorias = df['categoria'].unique()
matriz_cat = np.zeros((len(categorias), len(categorias)))

for i, cat1 in enumerate(categorias):
    for j, cat2 in enumerate(categorias):
        idx1 = df[df['categoria'] == cat1].index
        idx2 = df[df['categoria'] == cat2].index
        matriz_cat[i, j] = cosine_similarity(
            vectors[idx1].mean(axis=0).reshape(1, -1),
            vectors[idx2].mean(axis=0).reshape(1, -1)
        )[0][0]

# 📊 Visualização com heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(matriz_cat, xticklabels=categorias, yticklabels=categorias, annot=True, cmap="coolwarm")
plt.title("Similaridade média entre categorias")
plt.xlabel("Categoria Comparada")
plt.ylabel("Categoria Base")
plt.tight_layout()
plt.show()

```



```
# =====
# 🌀 PCA COM FILTRO POR CATEGORIA
# =====

def plot_pca_filtrado(vectors, labels, categorias, modo='2D', categoria_selecionada=None):
    if categoria_selecionada:
        mask = df['categoria'] == categoria_selecionada
        vectors = vectors[mask]
        labels = df['name'][mask]
        categorias = df['categoria'][mask]

    if modo == '2D':
        pca = PCA(n_components=2)
        coords = pca.fit_transform(vectors)

        plt.figure(figsize=(10, 8))
        plt.scatter(coords[:, 0], coords[:, 1], c='skyblue', s=50)
        for i, name in enumerate(labels):
            plt.text(coords[i, 0], coords[i, 1], name, fontsize=8)

        plt.title(f"Mapa 2D - Categoria: {categoria_selecionada or 'Todas'}")
        plt.xlabel("Componente 1")
        plt.ylabel("Componente 2")
        plt.tight_layout()
        plt.show()

    else:
        pca = PCA(n_components=3)
        coords = pca.fit_transform(vectors)

        fig = plt.figure(figsize=(10, 8))
        ax = fig.add_subplot(111, projection='3d')
        ax.scatter(coords[:, 0], coords[:, 1], coords[:, 2], c='skyblue', s=50)
        for i, name in enumerate(labels):
            ax.text(coords[i, 0], coords[i, 1], coords[i, 2], name, fontsize=8)

        ax.set_title(f"Mapa 3D - Categoria: {categoria_selecionada or 'Todas'}")
        ax.set_xlabel("Componente 1")
        ax.set_ylabel("Componente 2")
        ax.set_zlabel("Componente 3")
        plt.tight_layout()
        plt.show()

# Widgets de controle
modo_pca = widgets.ToggleButtons(
    options=['2D', '3D'],
    description='Visualização:',
    button_style='info'
)
```

```

categoria_dropdown = widgets.Dropdown(
    options=['Todas'] + sorted(df['categoria'].unique()),
    description='Categoria:',
    style={'description_width': 'initial'}
)

def atualizar_visualizacao(change=None):
    clear_output(wait=True)
    modo = modo_pca.value
    categoria = categoria_dropdown.value
    cat = None if categoria == 'Todas' else categoria
    plot_pca_filtrado(vectors, df['name'], df['categoria'], modo=modo, categoria_selecionada=cat)
    display(widgets.HBox([modo_pca, categoria_dropdown]))

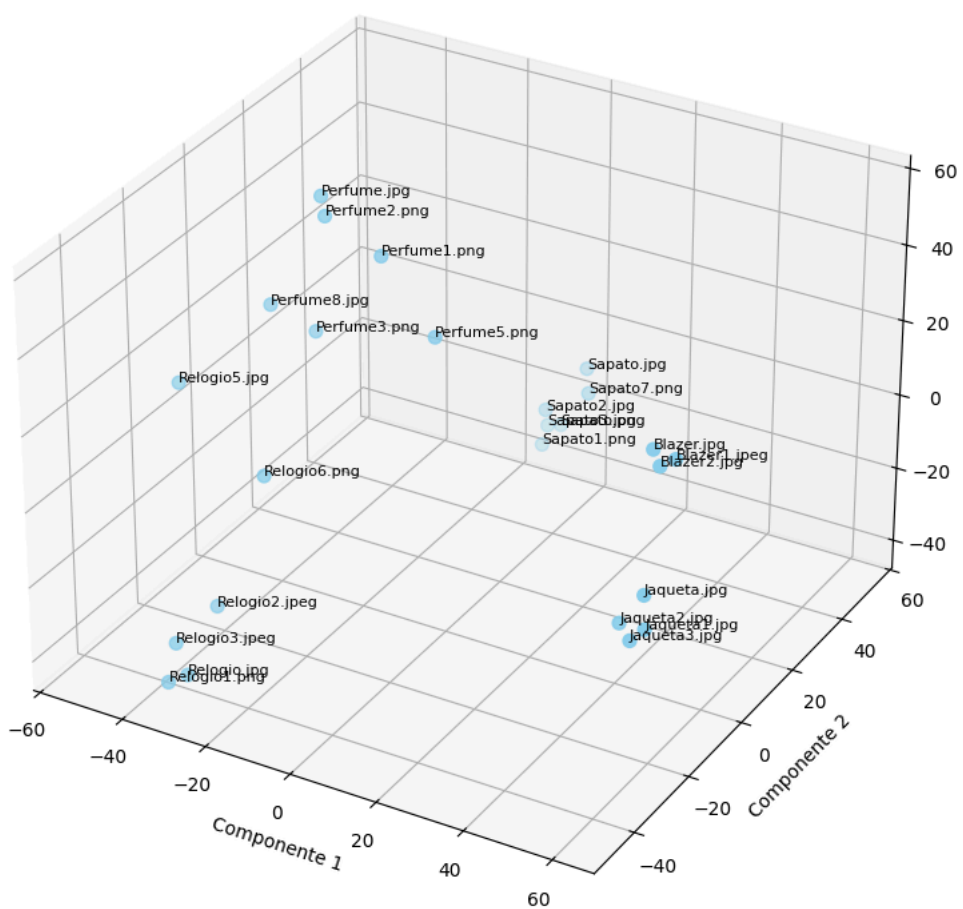
modo_pca.observe(atualizar_visualizacao, names='value')
categoria_dropdown.observe(atualizar_visualizacao, names='value')

# Exibe os controles e inicia com visualização padrão
atualizar_visualizacao()

```



Mapa 3D - Categoria: Todas



Visualização:

Categoria: