Content Recommendations System Report

This project implements a content-based recommendation system that suggests similar products based on their descriptions, tags, and metadata. The model uses TF-IDF vectorisation and cosine similarity to compute product similarity and generate recommendations.

Overview:

- Dataset: Walmart product reviews dataset (TSV format, ~5k records).
- Goal: Recommend products to users based on product descriptions, tags, brand, and categories.
- Approach: Preprocess dataset (handle missing values, select relevant columns). Convert text features into numerical vectors using TF-IDF. Compute cosine similarity between product vectors. Recommend the top N similar products for a given product.

Features:

Content-based filtering uses TF-IDF on product descriptions and tags, Cosine similarity for recommendation ranking, is Scalable to larger datasets, and a Modular notebook for experimentation

Tech Stack:

- Language: Python 3.x
- Libraries: Pandas, NumPy (data processing), Scikit-learn (TF-IDF, cosine similarity), Matplotlib, Seaborn (visualisation), SciPy (sparse matrix operations)

Code Implementation and Outcome:

Content Base Recommendation system (User Preferences or Items similarities)

[30]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity

tfidf_vectorizer = TfidfVectorizer(stop_words='english')
tfidf_matrix_content = tfidf_vectorizer.fit_transform(train_data['Tags'])
cosine_similarities_content = cosine_similarity(tfidf_matrix_content,tfidf_matrix_content)
```

[31]:
```
item_name = 'OPI Infinite Shine, Nail Lacquer Nail Polish, Bubble Bath'
item_index = train_data[train_data['Name']==item_name].index[0]
```

[32]:
```
similar_items = list(enumerate(cosine_similarities_content[item_index]))
```

[33]:
```
similar_items = sorted(similar_items, key=lambda x:x[1], reverse=True)
top_similar_items = similar_items[1:10]

recommended_items_indics = [x[0] for x in top_similar_items]
```

[34]:
```
train_data.iloc[recommended_items_indics][['Name','ReviewCount','Brand']]
```

[34]:

	Name	ReviewCount	Brand
156	OPI Nail Lacquer Polish .5oz/15mL - This Gown ...	0.0	opi
184	OPI Nail Gel Polish GelColor .5oz/15mL 3 CT Co...	0.0	opi
205	OPI Nail Lacquer - Dont Bossa Nova Me Around -...	0.0	opi
237	OPI Infinite Shine 2 Polish - ISL P33 - Alpaca...	5.0	opi
325	OPI Gel Polish Fall 2019 Scotland Collection G...	1.0	opi

[30]:
```
# Example: Get content-based recommendations for a specific item
item_name = 'Kokie Professional Matte Lipstick, Hot Berry, 0.14 fl oz'
content_based_rec = content_based_recommendations(train_data, item_name, top_n=8)

content_based_rec
```

[30]:

	Name	ReviewCount	Brand	ImageURL	Rating
3406	Kokie Professional Matte Lipstick, Firecracker...	0.0	kokie, cosmetics	https://i5.walmartimages.com/asr/8312221b-ed22...	0.0
546	Kokie Professional Matte Lipstick, Kiss Me, 0....	0.0	kokie, cosmetics	https://i5.walmartimages.com/asr/27dd82a2-2b9c...	0.0
2406	L.A. Colors Matte Lipstick, Tender Matte	3.0	colors	https://i5.walmartimages.com/asr/271264fb-e8c3...	3.7
4050	Kokie Professional Lip Poudre Liquid Matte Liq...	7.0	kokie, cosmetics	https://i5.walmartimages.com/asr/fdd7498c-319f...	3.4
4084	e.l.f. Mad for Matte 4 Piece Lip Color Set	0.0	cosmetics	https://i5.walmartimages.com/asr/e2d30304-edc9...	0.0
1559	LOreal Paris Colour Riche Matte Lip Liner, Mat...	495.0	paris	https://i5.walmartimages.com/asr/baf97085-7231...	4.4
2873	Kokie Professional Lip Poudre Liquid Matte Liq...	7.0	kokie, cosmetics	https://i5.walmartimages.com/asr/31c99d9b-ea11...	3.4
3023	Be Matte Lipstick - Pink	2.0	city, color	https://i5.walmartimages.com/asr/4425a13e-085f...	3.0

Conclusion:

This project demonstrates how a content-based recommendation system can be built using TF-IDF vectorisation and cosine similarity. By analysing product descriptions, tags, and categories, the system effectively suggests similar products to users. While the current implementation works well for static datasets, it can be further enhanced with hybrid methods, real-time personalisation, and deployment as a web application. The project highlights the potential of machine learning in improving customer experience through personalised recommendations, which are widely used in e-commerce platforms, streaming services, and digital marketplaces.