Meirkhanova Alma BDA-2105 Final Project Report

Introduction

Link for YOUTUBE video: https://youtu.be/NL-aqSnq7CU

Link for GITHUB: https://github.com/alma3003/final.git

Link for other solutions:

1. https://github.com/HasibAlMuzdadid/Recommender-Systems/blob/main/book%20recommender%20system/book%20recommender%20system.ipynb
2. https://github.com/aswintechguy/Machine-Learning-Projects/blob/master/Million%20Songs%20Dataset%20-%20Recommendation%20Engine/Million%20Songs%20Data%20-%20Recommendation%20Engine.ipynb
3. https://github.com/campusx-official/movie-recommender-system-tmdb-dataset/blob/main/notebook86c26b4f17.ipynb

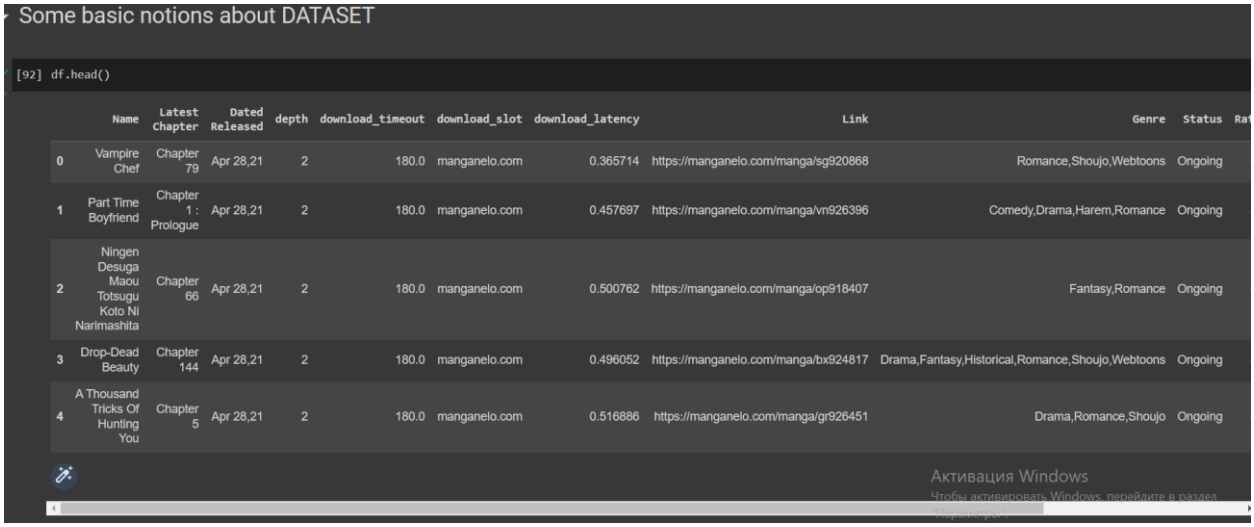
Problem: The problem was to create a recommendation system/engine for the Manga readers.

Current work: I created the recommendation system using the Manga dataset from Kaggle. I already uploaded it with comments. However, in short, I used the K-means and collaborative filtering, which can filter the data based on similar data. For that, I also used the cosine similarity. Analytically, it measures the cosine of the angle between two data points. The cosine similarity is advantageous because even if the two similar data points are far apart by the Euclidean distance (due to the size of the document), chances are they may still be oriented closer together. The smaller the angle, higher the cosine similarity. In the end, it recommends the top 5 manga. There is also summary and top 10 mangas by Rating.

Data and Methods: The data is MANGA dataset which includes columns like Name, Latest Chapter, Dated Released,depth,download_timeout,download_slot,download_latency,Link,Genre,Status,Rating,img-link. It is from KAGGLE:

https://www.kaggle.com/datasets/darknez/manga-dataset

Here some analysis and visualizations of dataset:



	Name	Latest Chapter	Dated Released	depth	download_timeout	download_slot	download_latency	Link	Genre	Status	Ra
0	Vampire Chef	Chapter 79	Apr 28,21	2	180.0	manganelo.com	0.365714	https://manganelo.com/manga/sg920868	Romance,Shoujo,Webtoons	Ongoing	
1	Part Time Boyfriend	Chapter 1 : Prologue	Apr 28,21	2	180.0	manganelo.com	0.457697	https://manganelo.com/manga/vn926396	Comedy,Drama,Harem,Romance	Ongoing	
2	Ningen Desuga Maou Totsugu Koto Ni Narimashita	Chapter 66	Apr 28,21	2	180.0	manganelo.com	0.500762	https://manganelo.com/manga/op918407	Fantasy,Romance	Ongoing	
3	Drop-Dead Beauty	Chapter 144	Apr 28,21	2	180.0	manganelo.com	0.496052	https://manganelo.com/manga/bx924817	Drama,Fantasy,Historical,Romance,Shoujo,Webtoons	Ongoing	
4	A Thousand Tricks Of Hunting You	Chapter 5	Apr 28,21	2	180.0	manganelo.com	0.516886	https://manganelo.com/manga/gr926451	Drama,Romance,Shoujo	Ongoing	

```
[94] print(f"Informations About Manga Dataset :\n")
     print(df.info())
```

```
Informations About Manga Dataset :

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14819 entries, 0 to 14818
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Name              14819 non-null  object
 1   Latest Chapter    14819 non-null  object
 2   Dated Released    14819 non-null  object
 3   depth             14819 non-null  int64
 4   download_timeout  14819 non-null  float64
 5   download_slot     14819 non-null  object
 6   download_latency  14819 non-null  float64
 7   Link              14819 non-null  object
 8   Genre             14643 non-null  object
 9   Status            14710 non-null  object
 10  Rating            14819 non-null  float64
 11  img-link          14819 non-null  object
dtypes: float64(3), int64(1), object(8)
memory usage: 1.4+ MB
```

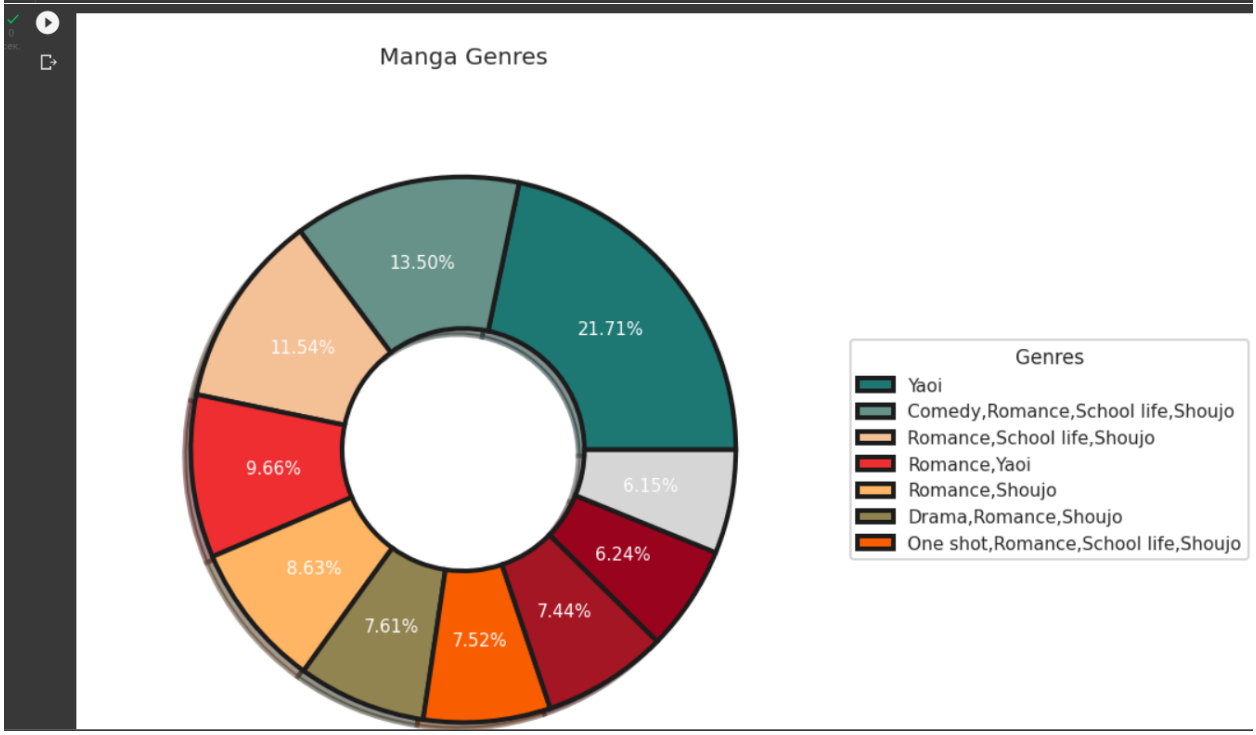
	Name	Latest Chapter	Dated Released	download_slot	Link	Genre	Status	img-link
count	14819	14819	14819	14819	14819	14643	14710	14819
unique	14786	7620	1200	1	14819	6677	29	14817
top	The	Chapter 1	Jan 20,16	manganelo.com	https://manganelo.com/manga/sg920868	Yaoi	Ongoing	https://avt.mkklcdnv6temp.com/44/h/5-158347362...
freq	23	291	4428	14819	1	255	9672	2

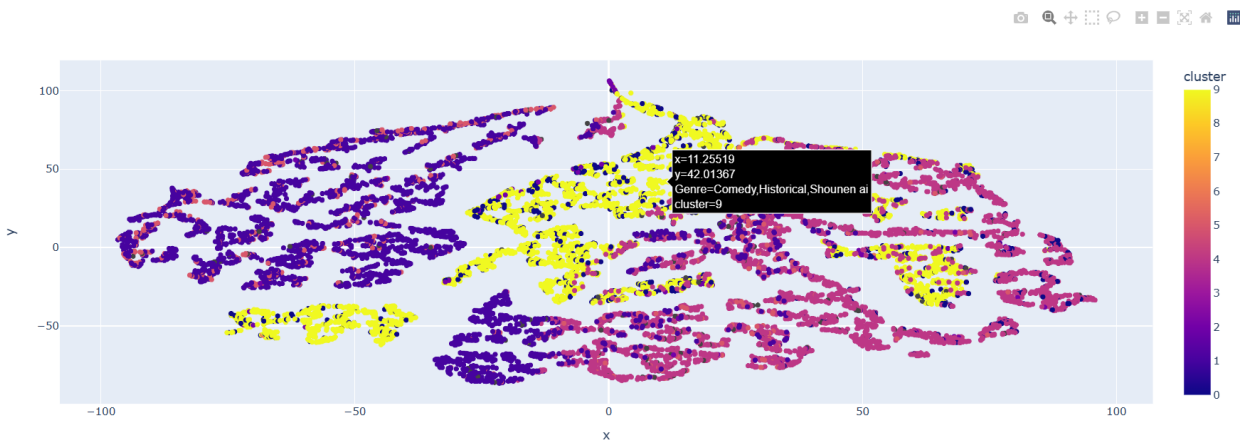
```
[ ] #the top manga genres
    print("Manga Genres :")
    top_manga_temp1["Genre"].value_counts().to_frame().T
```

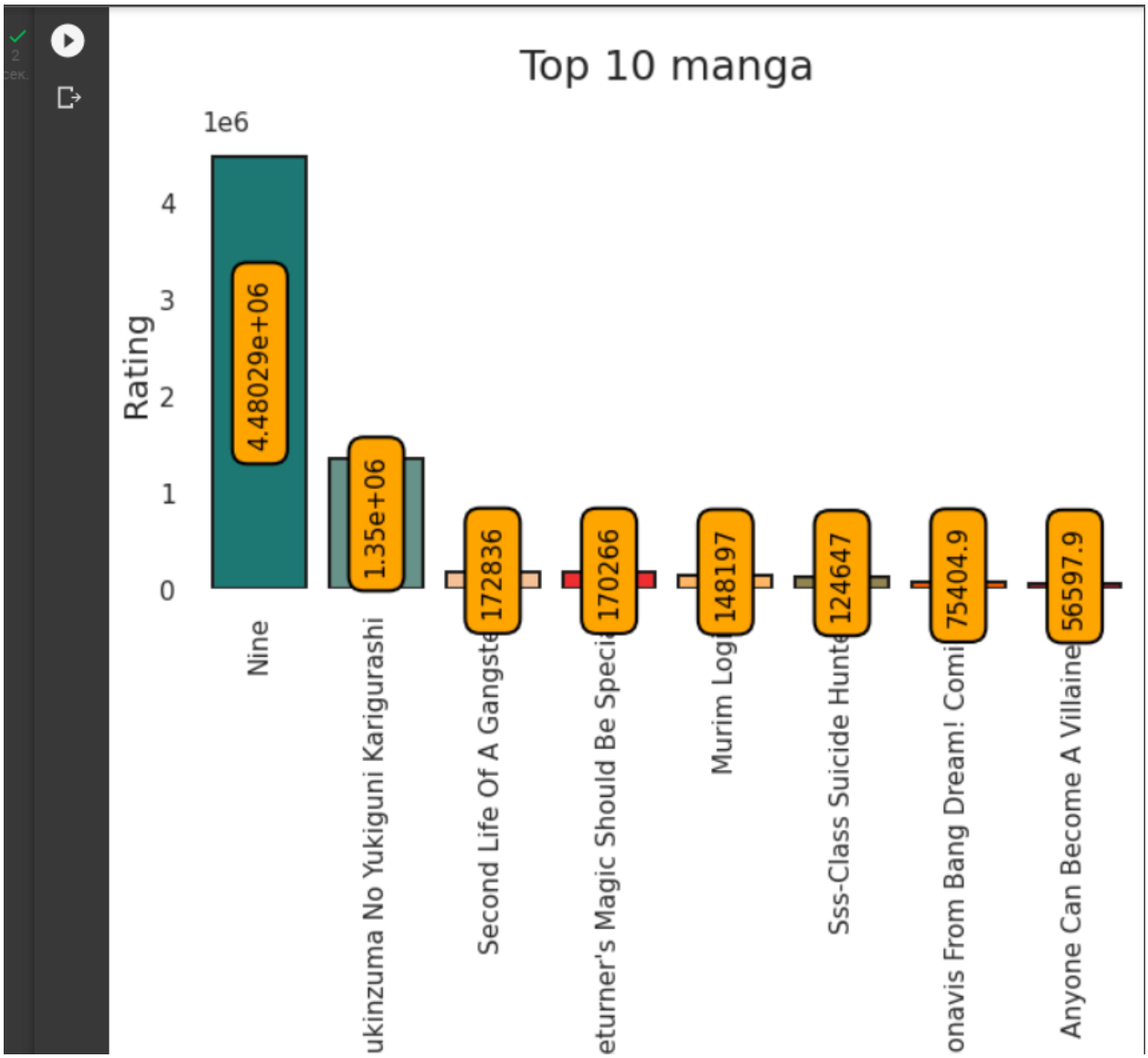
Manga Genres :

	Yaoi	Comedy,Romance,School life,Shoujo	Romance,School life,Shoujo	Romance,Yaoi	Romance,Shoujo	Drama,Romance,Shoujo	One shot,Romance,School life,Shoujo	Comedy,Drama,Romance,School life,Shoujo	Action,Adventure,Fantasy	Drama
Genre	254	158	135	113	101	89	88	87	73	

1 rows × 6632 columns







Description of the ML models you used with some theory: As I said, i used clustering, cosine similarity, preparation of data, k-means (nearest neighbor). In theory, i guess the cosine similarity is the best because it is not depend on size of the dataset, and still can be oriented closer.

Results:

Recommendations for Nettaigyo Wa Yuki Ni Kogareru readers :

No	Manga Name	Rating
1	Sexual Education 120%	4.360000
2	Nettaigyo Wa Yuki Ni Kogareru	4.350000
3	Futaribeya	4.380000
4	Sunami Yuuko And The Yuri People	4.660000
5	Jk Shousetsuka Ppoi!	3.750000

It is the one of the results of recommendation system. You can see there 5 top mangas that were recommended to user and it's ratings as well. In addition, I checked the result if these mangas were similar and they were. They have exactly the similar Genres.

Critical review of results:

```
69252481,"Comedy,School life,Seinen,Shoujo ai,Slice of life",Ongo
```

It is the Genre of Wa Yuki from the MANGA dataset. In addition, I consider the First recommended manga Sexual Education 120%.

```
493,"Comedy,School life,Seinen,Shoujo ai,Slice of life",Ongoing,4.3
```

As you can see, it is the same, thus it the top recommended manga.

Links and codes that I used during the project work:

1. https://www.kaggle.com/code/hasibalmuzdadid/anime-ratings-analysis-recommender-system/notebook
2. https://www.kaggle.com/datasets/darknez/manga-dataset
3. https://www.kaggle.com/code/tj00001/building-music-recommendation-system-using-spotify