

Cloud Assignment

Outputs:

Data:

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...	ratings_count	work_ratings_count	work_text_r
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	...	4780653	4942365	
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...	4602479	4800065	
2	3	41865	41865	3212258	226	316015849	9.780316e+12	Stephenie Meyer	2005.0	Twilight	...	3866839	3916824	
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	...	2346404	2478609	
4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	...	1903563	2216814	

Number of missing values:

```
Total number of missing values: 260
Percentage of missing values for training data:
book_id                0.000000
goodreads_book_id      0.000000
best_book_id           0.000000
work_id                0.000000
books_count            0.000000
isbn                   3.840473
isbn13                 3.249631
authors                0.000000
original_publication_year 0.221566
original_title          3.840473
title                  0.000000
language_code           8.050222
average_rating          0.000000
ratings_count           0.000000
work_ratings_count      0.000000
work_text_reviews_count 0.000000
ratings_1               0.000000
ratings_2               0.000000
ratings_3               0.000000
ratings_4               0.000000
ratings_5               0.000000
image_url               0.000000
small_image_url         0.000000
dtype: float64
The column with the highest missing values percentage is: language_code
```

Number of missing values after cleaning data:

```
Total number of missing values: 0
C:\Users\tenovo\AppData\Local\Temp\ipykernel_24024\4119370490.py:4: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance
categorical_cols = [col for col in data.columns if pd.api.types.is_categorical_dtype(data[col]) or data[col].dtype == 'object']
```

Data that contain only harry potter books:

	book_id	goodreads_book_id	best_book_id	work_id	books_count \
1	2.0	3.0	3.0	4640799.0	491.0
6	18.0	5.0	5.0	2402163.0	376.0
8	21.0	2.0	2.0	2809203.0	307.0
9	23.0	15881.0	15881.0	6231171.0	398.0
10	24.0	6.0	6.0	3046572.0	332.0
11	25.0	136251.0	136251.0	2963218.0	263.0
12	27.0	1.0	1.0	41335427.0	275.0
96	422.0	862041.0	862041.0	2962492.0	76.0
613	3753.0	10.0	10.0	21457570.0	6.0
1036	7018.0	483445.0	483445.0	471792.0	42.0
1266	9048.0	2002.0	2002.0	8621948.0	5.0

	original_publication_year \
1	1997.0
6	1999.0
8	2003.0
9	1998.0
10	2000.0
11	2007.0
12	2005.0
96	1998.0
613	2005.0
1036	2001.0
1266	2001.0

	work_ratings_count	work_text_reviews_count	ratings_1	ratings_2 \
1	4800065.0	75867.0	75504.0	101676.0
6	1969375.0	36099.0	6716.0	20413.0
8	1840548.0	28685.0	9528.0	31577.0
9	1906199.0	34172.0	8253.0	42251.0
10	1868642.0	31084.0	6676.0	20210.0
11	1847395.0	51942.0	9363.0	22245.0
12	1785676.0	27520.0	7308.0	21516.0
96	204125.0	6508.0	1105.0	1285.0
613	26274.0	882.0	203.0	186.0
1036	15145.0	267.0	329.0	1125.0
1266	11732.0	185.0	106.0	304.0

	isbn	isbn13	authors \
1	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré
6	043965548X	9.780440e+12	J.K. Rowling, Mary GrandPré, Rufus Beck
8	439358078	9.780439e+12	J.K. Rowling, Mary GrandPré
9	439064864	9.780439e+12	J.K. Rowling, Mary GrandPré
10	439139600	9.780439e+12	J.K. Rowling, Mary GrandPré
11	545010225	9.780545e+12	J.K. Rowling, Mary GrandPré
12	439785960	9.780440e+12	J.K. Rowling, Mary GrandPré
96	545044251	9.780545e+12	J.K. Rowling
613	439827604	9.780440e+12	J.K. Rowling
1036	042519891X	9.780425e+12	David Colbert
1266	043932162X	9.780439e+12	J.K. Rowling

	original_title	... ratings_count \
1	Harry Potter and the Philosopher's Stone	... 4602479.0
6	Harry Potter and the Prisoner of Azkaban	... 1832823.0
8	Harry Potter and the Order of the Phoenix	... 1735368.0
9	Harry Potter and the Chamber of Secrets	... 1779331.0
10	Harry Potter and the Goblet of Fire	... 1753043.0
11	Harry Potter and the Deathly Hallows	... 1746574.0
12	Harry Potter and the Half-Blood Prince	... 1678823.0
96	Complete Harry Potter Boxed Set	... 190050.0
613	Harry Potter Collection (Harry Potter, #1-6)	... 24618.0
1036	The Magical Worlds of Harry Potter: A Treasury...	... 13820.0
1266	Dark Reunion	... 10736.0

	ratings_3	ratings_4	ratings_5 \
1	455024.0	1156318.0	3011543.0
6	166129.0	509447.0	1266670.0
8	180210.0	494427.0	1124806.0
9	242345.0	548266.0	1065084.0
10	151785.0	494926.0	1195045.0
11	113646.0	383914.0	1318227.0
12	136333.0	459028.0	1161491.0
96	7020.0	30666.0	164049.0
613	946.0	3891.0	21048.0
1036	3766.0	3593.0	6332.0
1266	1548.0	2595.0	7179.0

	image_url \
1	https://images.gr-assets.com/books/1474154022m...
6	https://images.gr-assets.com/books/1499277281m...
8	https://images.gr-assets.com/books/1387141547m...
9	https://images.gr-assets.com/books/1474169725m...
10	https://images.gr-assets.com/books/1361482611m...
11	https://images.gr-assets.com/books/1474171184m...
12	https://images.gr-assets.com/books/1361039191m...
96	https://images.gr-assets.com/books/1392579059m...
613	https://images.gr-assets.com/books/1328867351m...
1036	https://s.gr-assets.com/assets/nophoto/book/11...
1266	https://images.gr-assets.com/books/1479775271m...

	small_image_url
1	https://images.gr-assets.com/books/1474154022s...
6	https://images.gr-assets.com/books/1499277281s...
8	https://images.gr-assets.com/books/1387141547s...
9	https://images.gr-assets.com/books/1474169725s...
10	https://images.gr-assets.com/books/1361482611s...
11	https://images.gr-assets.com/books/1474171184s...
12	https://images.gr-assets.com/books/1361039191s...
96	https://images.gr-assets.com/books/1392579059s...
613	https://images.gr-assets.com/books/1328867351s...
1036	https://s.gr-assets.com/assets/nophoto/book/50...
1266	https://images.gr-assets.com/books/1479775271s...

Most selling harry potter book:

```
The most selling harry potter book: 1    Harry Potter and the Philosopher's Stone  
Name: original_title, dtype: object
```

Average rating of harry potter books:

```
Average rating harry potter: 4.482727272727273
```