

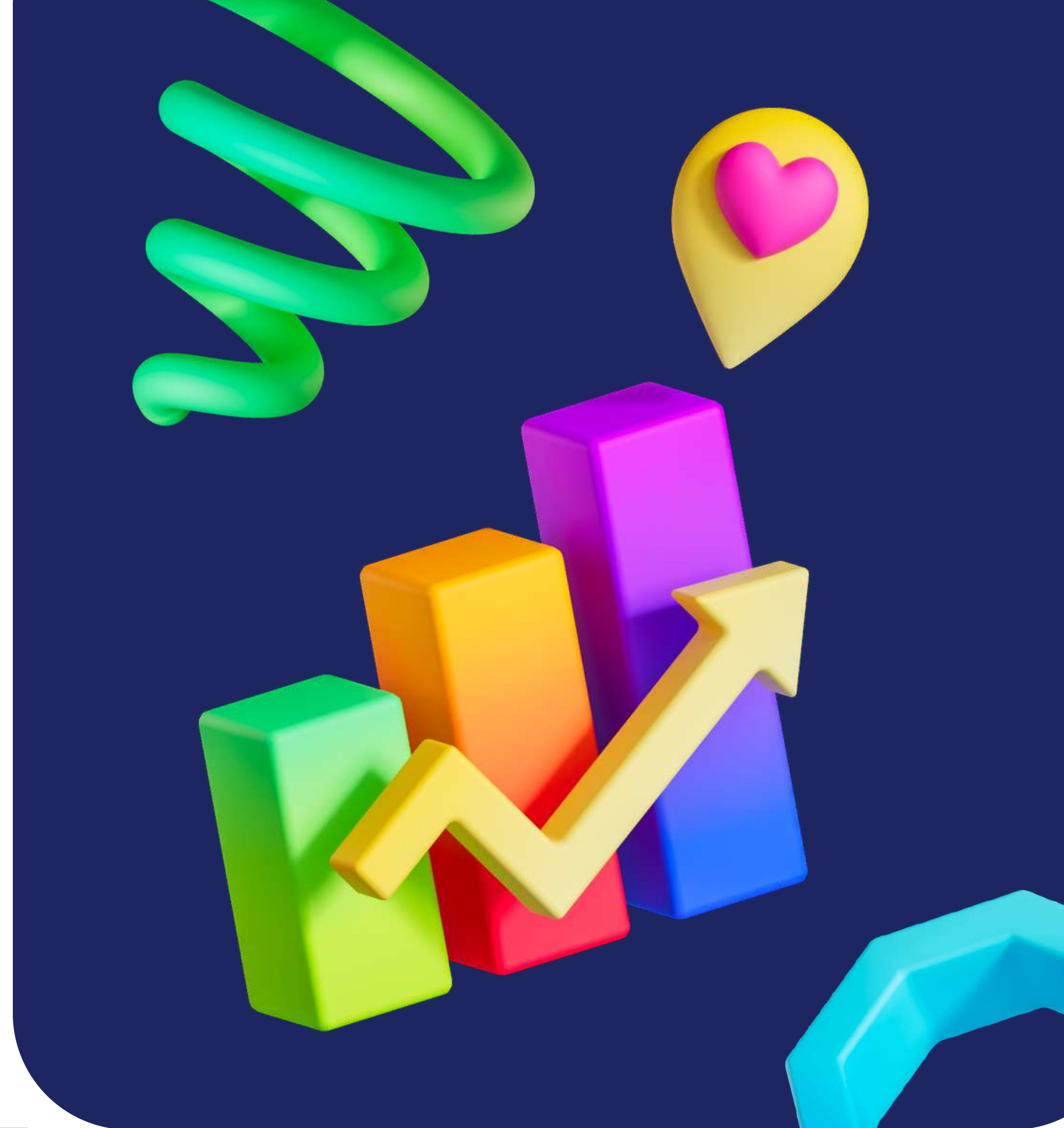
# Korean Drama Recommendation Using Python

FRIDA PUTRIASSA - 2110512104

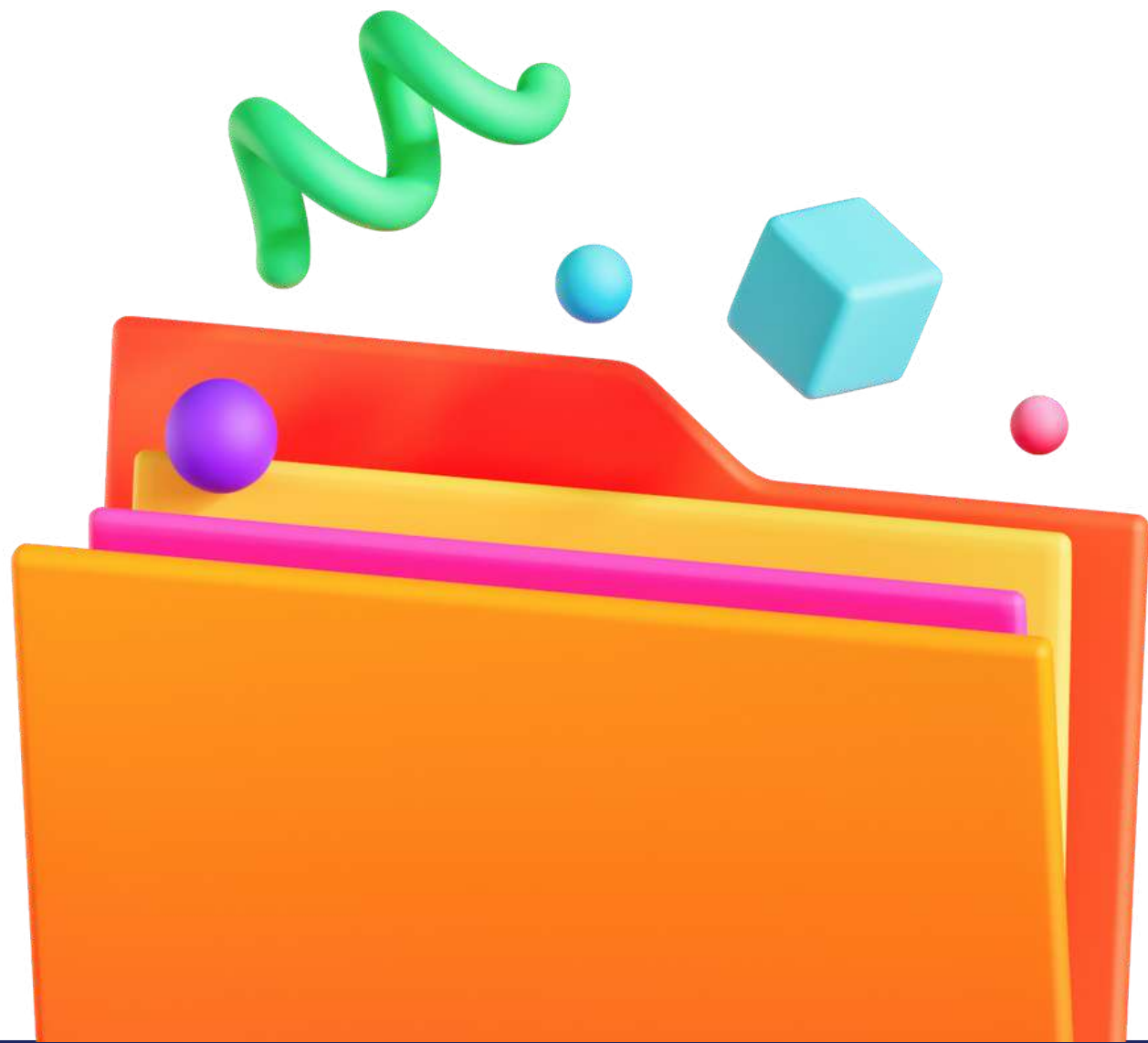


Portfolio Project  
Bisa.ai

Course :  
Data Science



# AGENDA



3

Introduction

---

4

Dataset Explanation

---

5

Importing Library

---

6

Data Preparation

---

7

Data Visualization

---

8

Conclusion

---



# Introduction

The 2023 Korean Drama Recommendation System utilizes machine learning to suggest Korean dramas from that year. It considers three key factors: duration, episode count, and overall reviews. By analyzing these factors, the system provides personalized recommendations, simplifying the selection process for viewers. Whether you prefer shorter dramas, longer storylines, or highly-rated series, this system offers tailored suggestions to enhance your Korean drama viewing experience.

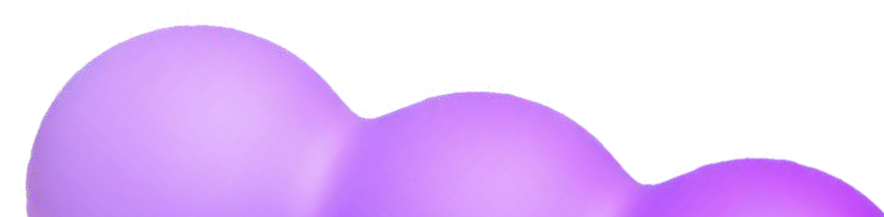






# DATASET EXPLANATION

Korean Drama from 2015-2023 with Actors & Reviews



This dataset particularly focusing on the popular Korean Drama website, MyDramaList. The dataset I've compiled is a spinoff of the "Top 100 Korean Drama (MyDramaList)" dataset.

# Dataset Contents

Link Dataset : [www.kaggle.com/datasets/chanoncharuchinda/top-100-korean-drama-mydramalists](https://www.kaggle.com/datasets/chanoncharuchinda/top-100-korean-drama-mydramalists)



## korean\_drama.csv

This file contains information about 1,752 Korean dramas that can be explored. It encompasses details about these dramas.

## recommendations.csv

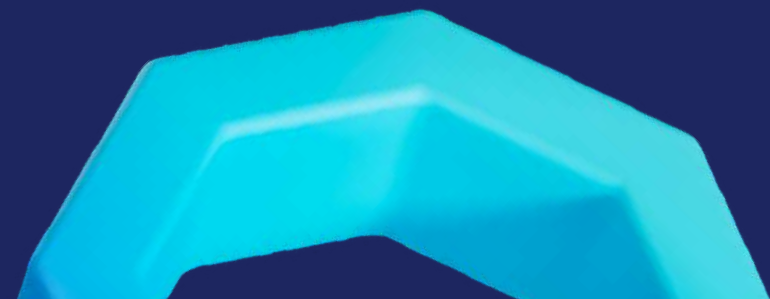
This file contains recommendations that have been given to various dramas that are a result of contributions from the MyDramaList community.

## review.csv

This file contains user reviews of the dramas, offering insights into what viewers think about the shows.

## wiki\_actor.csv

This dataset provides information about actors, primarily the dramas they've been a part of and the roles they've played.



In this recommendation system, we only use two table, which is korean\_drama and reviews.

Table : Korean\_Drama

Memiliki 17 kolom, yaitu

- kdrama\_id
- drama\_name
- year
- director
- screenwriter
- country
- type
- tot\_eps
- duration
- start\_dt
- end\_dt
- aired\_on
- org\_net
- content\_rt
- synopsis
- rank
- pop

Table : Korean\_Drama

Memiliki 10 kolom, yaitu

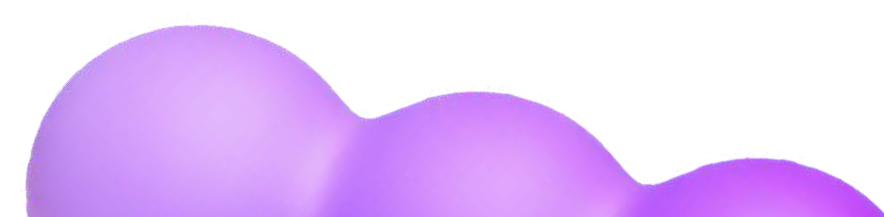
- user\_id
- title
- story\_score
- acting\_cast\_score
- music\_score
- rewatch\_value\_score
- overall\_score
- review\_text
- ep\_watched
- n\_helpful





# IMPORTING LIBRARY

Korean Drama from 2015-2023 with Actors & Reviews



# Library

- **NumPy**: Python library for numerical operations with arrays.
- **Pandas**: Python library for data manipulation and analysis, ideal for tables.
- **Matplotlib**: Python library for creating various types of data visualizations.

```
import numpy as np  
import pandas as pd
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

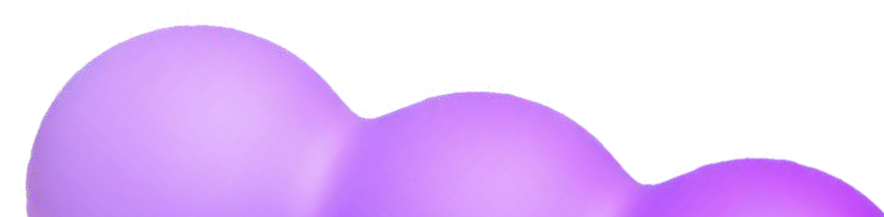






# DATA PREPARATION

Korean Drama from 2015-2023 with Actors & Reviews



Importing csv file into dataframe (csv korean\_drama into df\_kdrama) and viewing df\_kdrama.

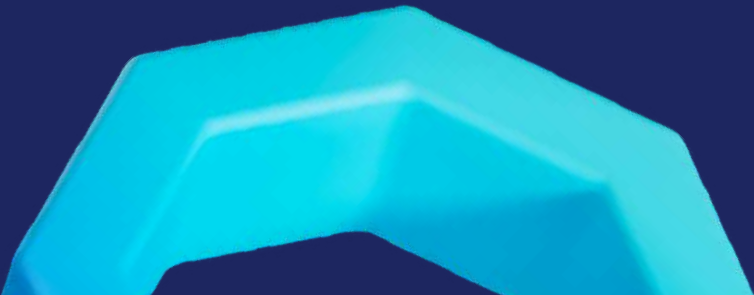
✓  
0s

[4] df\_kdrama

✓  
0s

[4] df\_kdrama

kdrama_id	drama_name	year	director	screenwriter	country	type	tot_eps	duration	start_dt	end_dt	aired_on	org_net	content_rt	synopsis
i73ab11e9e8	Sing My Crush	2023	['So Joon Moon']	NaN	South Korea	Drama	8	1500.0	Aug 2, 2023	Aug 2, 2023	Wednesday	NaN	Not Yet Rated	Follow the story of acquaintances Ba Ram and H...
y129984274c	D.P. Season 2	2023	NaN	['Kim Bo Tong']	South Korea	Drama	6	3000.0	Jul 28, 2023	Jul 28, 2023	Friday	Netflix	15+ - Teens 15 or older	This unfolding story ensues when military dese...
cee6f16c9fdf	Shadow Detective Season 2	2023	['Han Dong Hwa']	['Song Jung Woo', 'Hwang Seol Hun']	South Korea	Drama	8	3300.0	2023-07-05	2023-07-26	Wednesday	Disney+ Hulu	15+ - Teens 15 or older	Unfolds the ultimate counterattack of veteran ...
f5a7a53a328	To Be Honest	2023	NaN	NaN	South Korea	Drama	3	600.0	2023-06-30	2023-07-14	Friday	NaN	Not Yet Rated	Don't you have those days where the whole univ...



Finding missing value and checking each column for missing values

✓  
0s

```
print(df_kdrama.isna())
```

	kdrama_id	drama_name	year	director	screenwriter	country	type	\
0	False	False	False	False	True	False	False	
1	False	False	False	True	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	False	True	True	False	False	
4	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	
1747	False	False	False	False	False	False	False	
1748	False	False	False	False	False	False	False	
1749	False	False	False	False	False	False	False	
1750	False	False	False	False	False	False	False	
1751	False	False	False	True	True	False	False	

	tot_eps	duration	start_dt	end_dt	aired_on	org_net	content_rt	\
0	False	False	False	False	False	True	False	
1	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	False	False	False	True	False	
4	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	
1747	False	False	False	False	False	False	False	
1748	False	False	False	False	False	False	False	
1749	False	False	False	False	False	False	False	

✓  
0s

```
[9] print(df_kdrama.isna().sum())
```

kdrama_id	0
drama_name	0
year	0
director	716
screenwriter	793
country	0
type	0
tot_eps	0
duration	24
start_dt	0
end_dt	0
aired_on	232
org_net	408
content_rt	0
synopsis	168
rank	0
pop	0
dtype: int64	



## Replacing missing values and understanding data

```
df_kdrama.fillna(0)
```

	kdrama_id	drama_name	year	dire
0	661d4193916c4e71a2c70473ab11e9e8	Sing My Crush	2023	['So M
1	5ffcbeaa17114714af1959129984274c	D.P. Season 2	2023	
2	65075cb9c1a54be4a441cee6f16c9fdf	Shadow Detective Season 2	2023	
3	df0f0ac4b3ff4b15afa26f5a7a53a328	To Be Honest	2023	

```
df_kdrama.describe()
```

	year	tot_eps	duration	rank	pop
count	1752.000000	1752.000000	1728.000000	1752.000000	1752.000000
mean	2019.006849	18.996005	2149.791667	22247.869292	22153.746575
std	2.317455	25.618394	1532.133619	27688.839480	37776.993814
min	2015.000000	1.000000	60.000000	9.000000	1.000000
25%	2017.000000	8.000000	720.000000	2441.500000	915.750000
50%	2019.000000	12.000000	1800.000000	6265.500000	3698.500000
75%	2021.000000	16.000000	3600.000000	49204.500000	12086.250000
max	2023.000000	150.000000	9180.000000	99999.000000	99999.000000

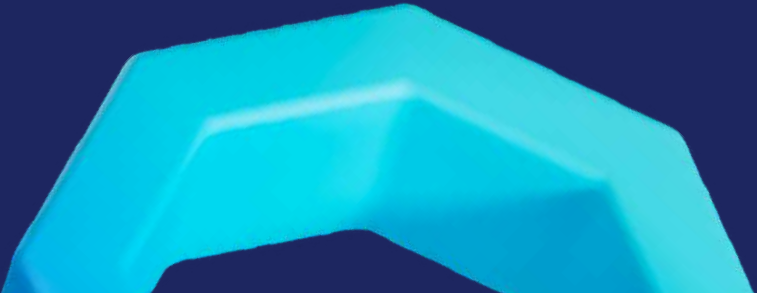
Import csv data reviews into dataframe  
and checking missing values

```
# Import data reviews
df_review= pd.read_csv('/content/KDrama/reviews.csv')
```

df_review									
	user_id	title	story_score	acting_cast_score	music_score	rewatch_value_score	overall_score	review_text	ep_w
	3eeb54b265598dbbb55	Sing My Crush	9.0	9.0	10.0	9.0	9.0	the Best Song from my Crush This will now ran...	e
	3eeb54b265598dbbb55	Happy Merry Ending	5.0	7.0	9.0	4.0	6.5	I'm Happy and Merry That it Ended Quickly The ...	e
	3eeb54b265598dbbb55	Duty After School: Part 2	4.0	9.0	3.0	1.0	4.0	This PART 2 Should Never Have Happened If you...	e
	3eeb54b265598dbbb55	Our Dating Sim	9.0	9.5	9.0	9.0	9.0	I want to play this dating sim! I want to pla...	e
		The							

```
# checking missing values tiap kolom
print(df_review.isna().sum())
```

```
user_id      0
title        0
story_score   0
acting_cast_score  0
music_score   0
rewatch_value_score  0
overall_score  0
review_text   6
ep_watched    0
n_helpful     0
dtype: int64
```



Link Dataset : [www.kaggle.com/datasets/chanoncharuchinda/top-100-korean-drama-mydramalists](https://www.kaggle.com/datasets/chanoncharuchinda/top-100-korean-drama-mydramalists)



```
# checking missing values tiap kolom  
print(df_review.isna().sum())
```

```
user_id          0  
title            0  
story_score      0  
acting_cast_score 0  
music_score      0  
rewatch_value_score 0  
overall_score    0  
review_text      6  
ep_watched       0  
n_helpful        0  
dtype: int64
```

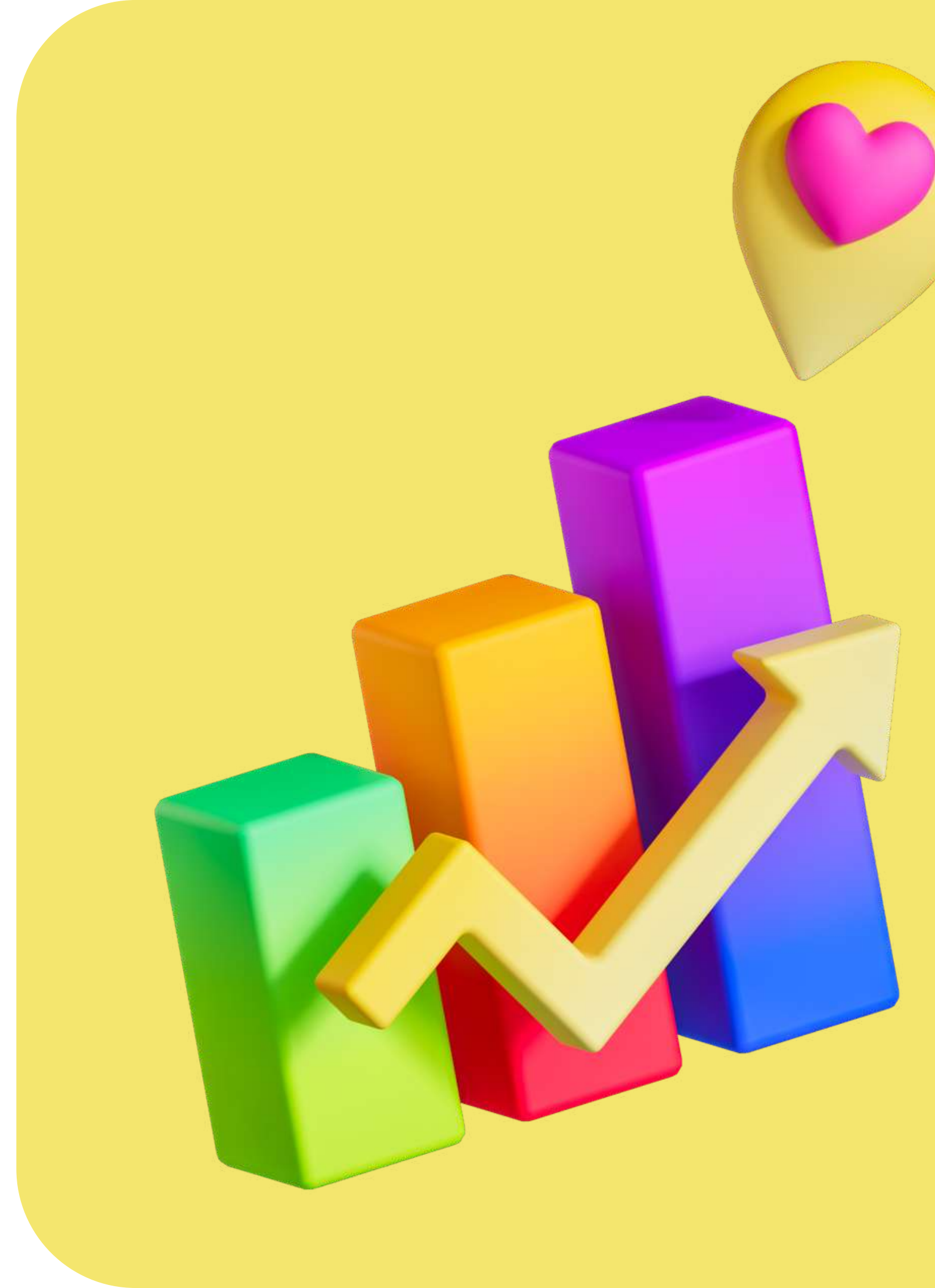
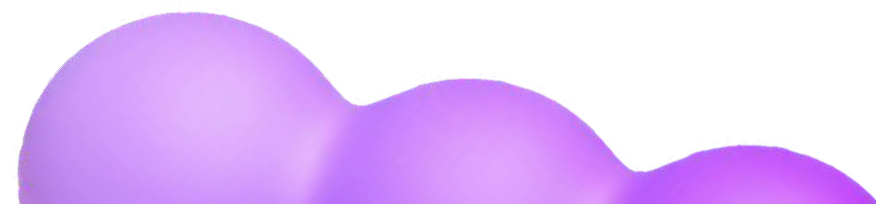






# DATA VISUALIZATION

Korean Drama from 2015-2023 with Actors & Reviews

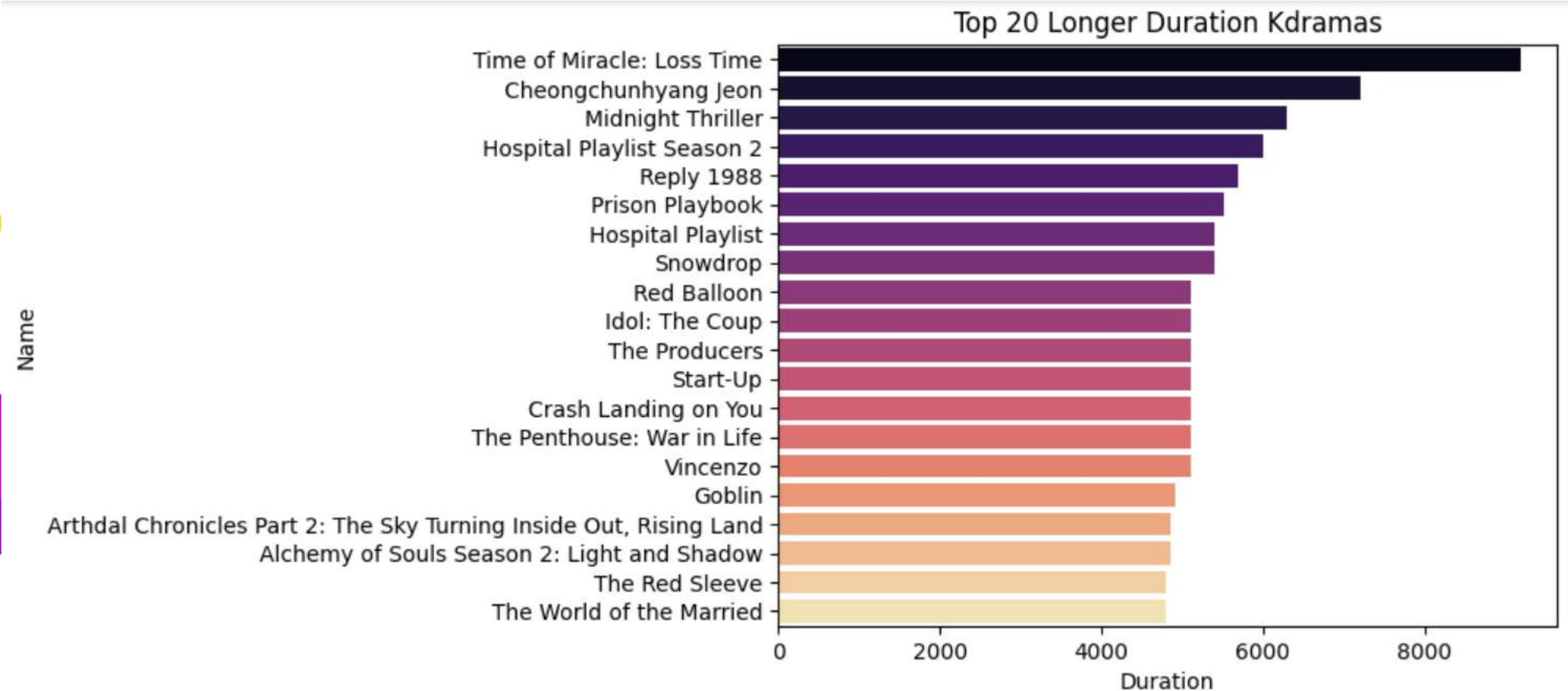


# TOP 20 LONGER DURATION KOREAN DRAMAS

For a Korean drama enthusiast accustomed to watching longer-duration dramas, they may not be disappointed with the duration of the following Korean drama. Topping the list as the Korean drama with the highest number of duration is "Time of Miracle: Loss Time"

```
# Top 20 Longer Duration
drama_sorted = df_kdrama.sort_values("duration", ascending = False)
drama20 = drama_sorted.iloc[0:20]
sns.barplot(data = drama20, x = "duration", y = "drama_name", palette = "magma")

plt.title("Top 20 Longer Duration Kdramas")
plt.ylabel("Name")
plt.xlabel("Duration")
plt.show()
```



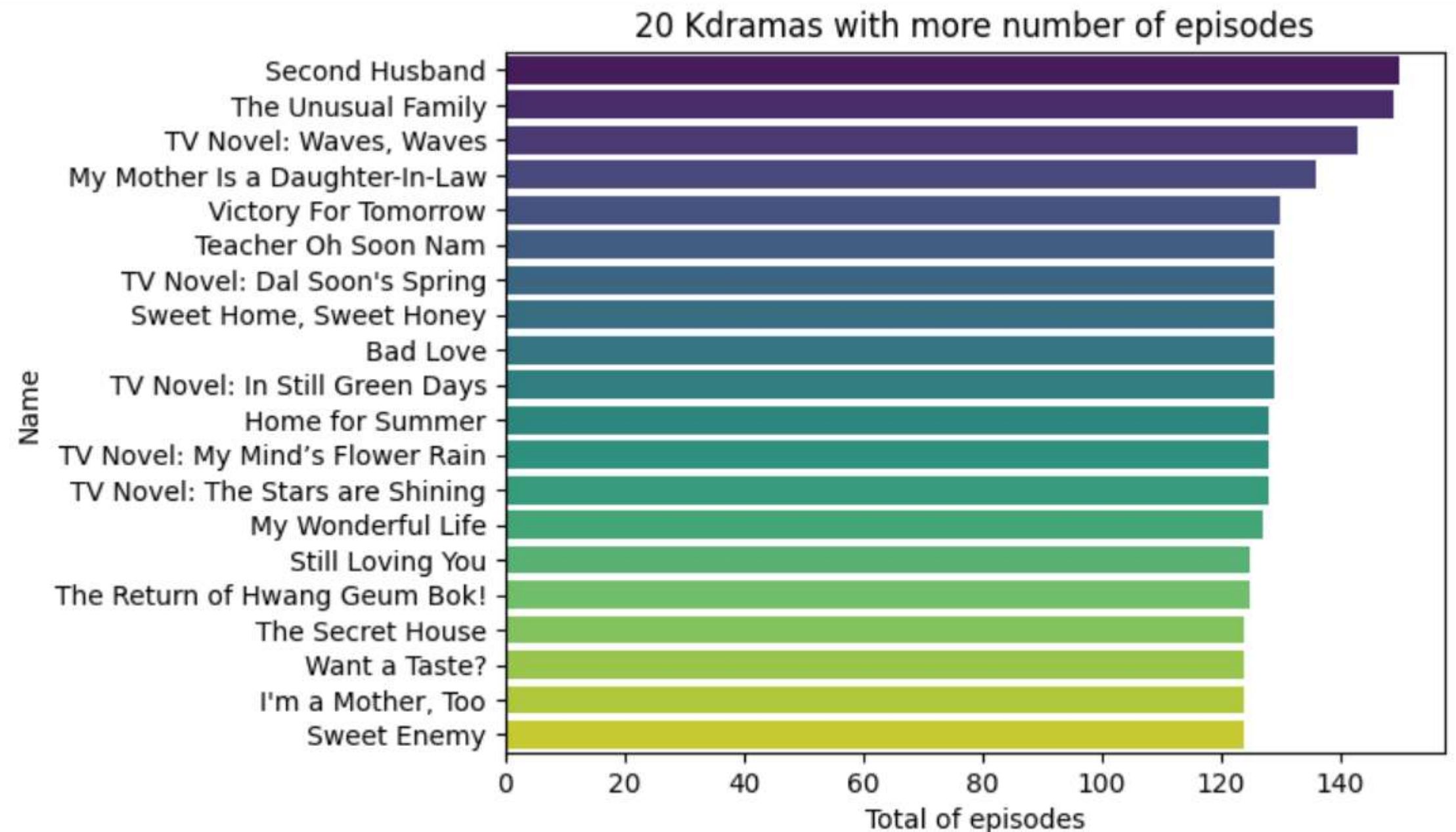


# 20 KOREAN DRAMA WITH MORE NUMBER OF EPISODES

For a Korean drama enthusiast accustomed to watching plenty episodes dramas, they may not be disappointed with the episodes of the following Korean drama. Topping the list as the Korean drama with the highest number of episodes is "Second Husband."

```
# Top 20 Episodes Count
episodes_sorted = df_kdrama.sort_values("tot_eps", ascending = False)
episodes = episodes_sorted.iloc[0:20]
sns.barplot(data = episodes, x = "tot_eps", y = "drama_name", palette = "viridis")

plt.title("20 Kdramas with more number of episodes")
plt.ylabel("Name")
plt.xlabel("Total of episodes")
plt.show()
```





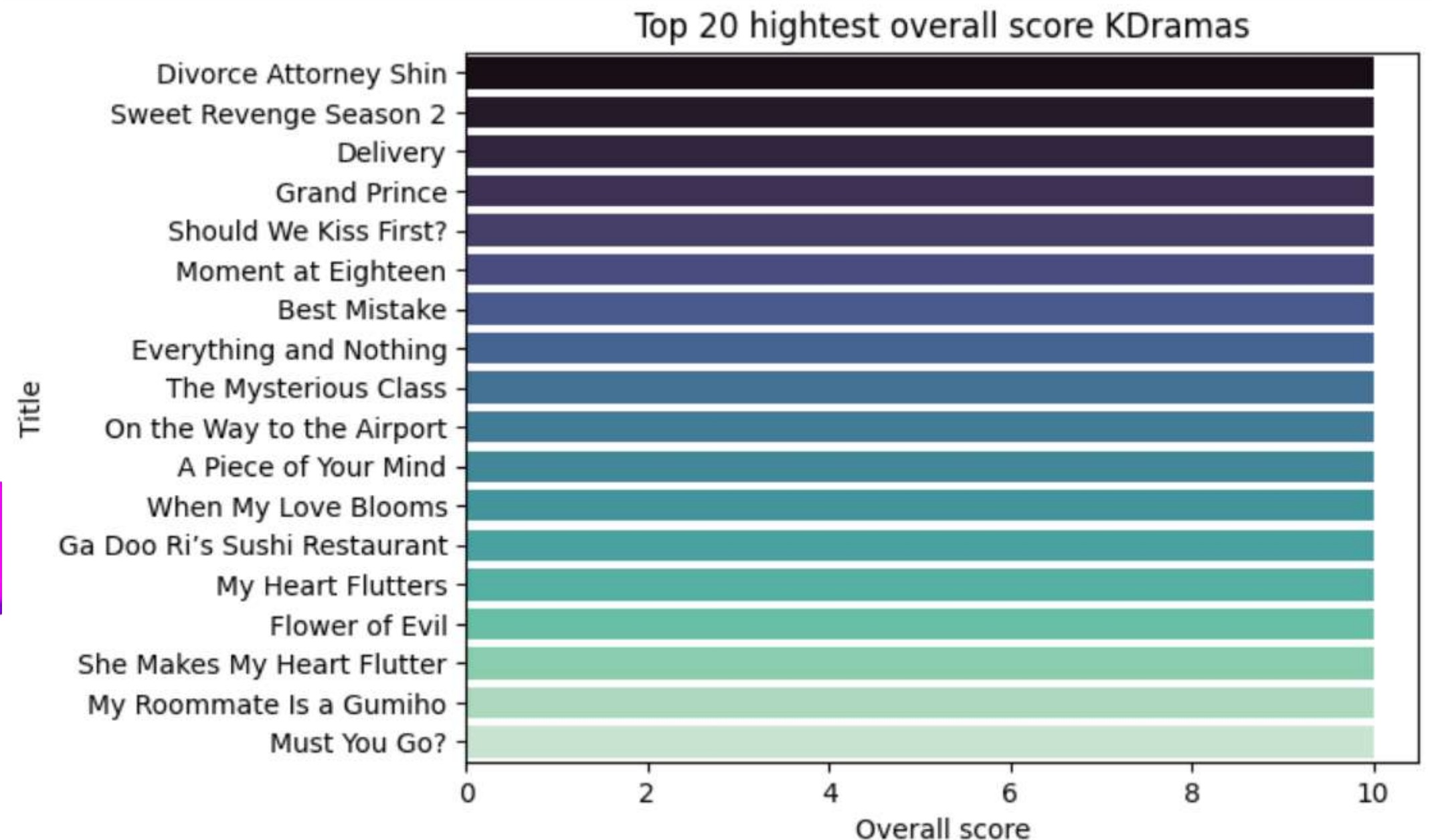
# TOP 20 HIGHEST OVERALL SCORE KOREAN DRAMAS

For the one who first time watching korean drama, or for korean drama addicted, both can add some of this drama title to your watching list.

```
# Top 20 highest overall score KDramas
```

```
overall_sorted = df_review.sort_values("overall_score", ascending= False)  
overall = overall_sorted.iloc[0:20]
```

```
sns.barplot(data = overall, x = "overall_score", y = "title", palette = "mako")  
plt.xlabel("Overall score")  
plt.ylabel("Title")  
plt.title("Top 20 highest overall score KDramas")
```



# AMOUNT OF KOREAN DRAMAS (2015 - 2023)

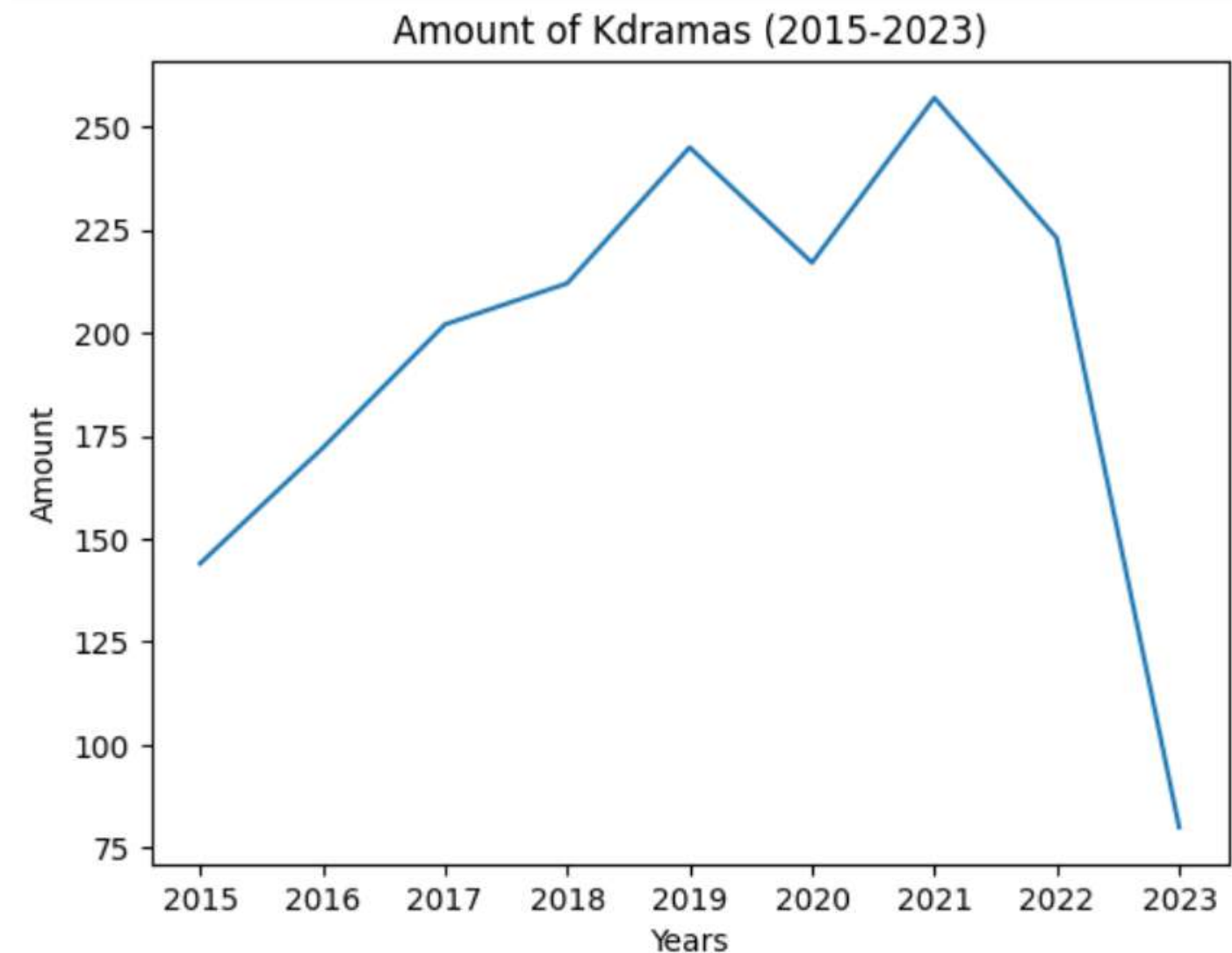
It can be observed that from 2015 to 2017, the number of Korean dramas consistently increased. However, in the following years, there was a fluctuation in the trend. Notably, from 2022 to 2023, there is a significant drop. This could be attributed to the fact that this data was collected in mid-2023, and it may not account for all the dramas released in that year.

```
# KDrama Amount from 2015 until 2023


amount= df_kdrama["year"].value_counts()
print(amount)
sns.lineplot(data= amount)
sns.set_style(rc = {'axes.facecolor': 'lightsteelblue'})
plt.xlabel("Years")
plt.ylabel("Amount")
plt.title("Amount of Kdramas (2015-2023)")
```

2021	257
2019	245
2022	223
2020	217
2018	212
2017	202
2016	172
2015	144
2023	80

Name: year, dtype: int64  
Text(0.5, 1.0, 'Amount of

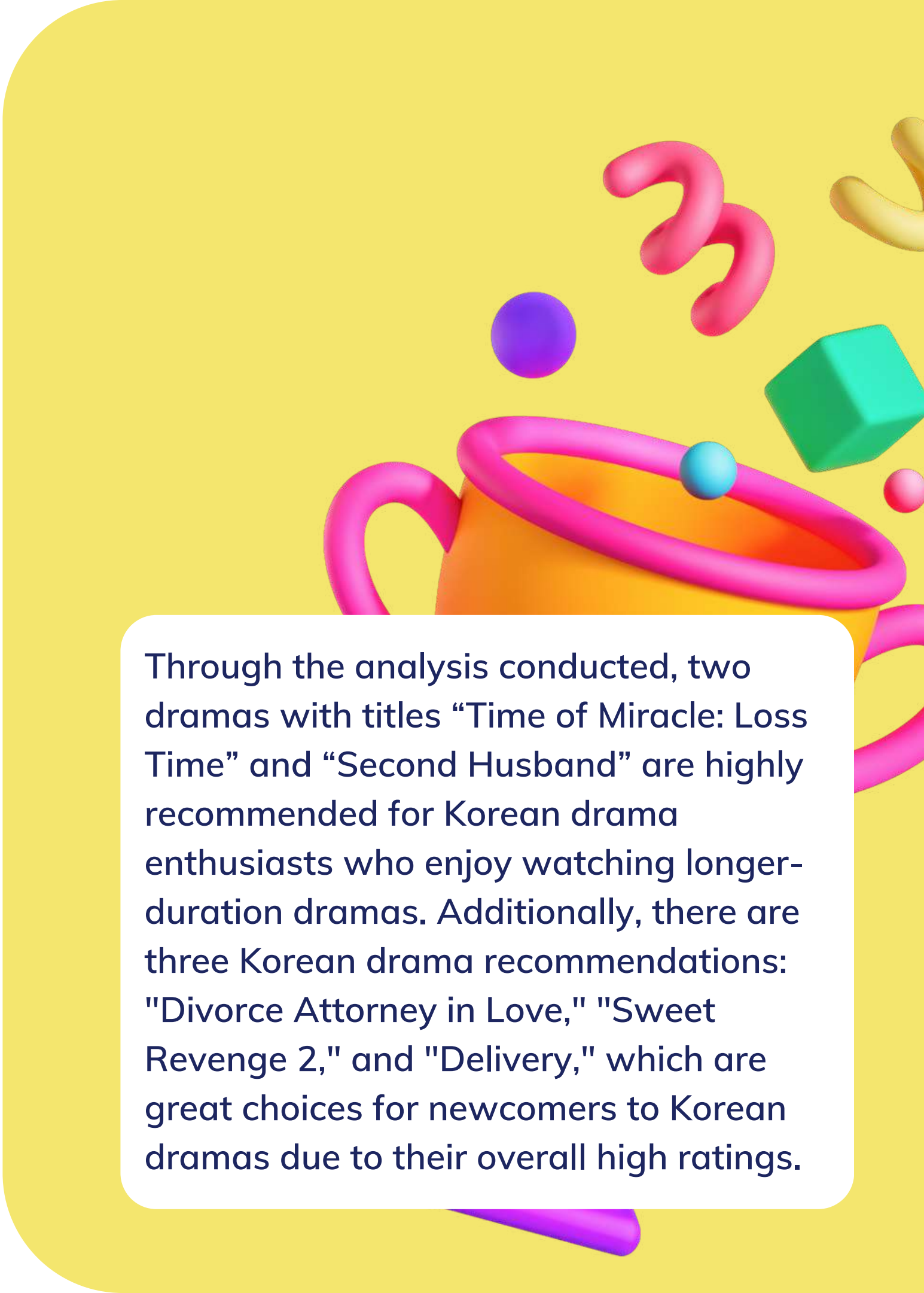






# CONCLUSION

Korean Drama from 2015-2023 with Actors & Reviews



Through the analysis conducted, two dramas with titles “Time of Miracle: Loss Time” and “Second Husband” are highly recommended for Korean drama enthusiasts who enjoy watching longer-duration dramas. Additionally, there are three Korean drama recommendations: “Divorce Attorney in Love,” “Sweet Revenge 2,” and “Delivery,” which are great choices for newcomers to Korean dramas due to their overall high ratings.