

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ПРИКАРПАТСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ВАСИЛЯ СТЕФАНІКА  
Кафедра комп'ютерних наук та інформаційних систем

**Індивідуальний проект**  
з курсу Python for Data Science  
на тему:  
«Аналіз датасету 52,000 Animation Movie Details»

**Виконала:**

студентка групи КН-41

Бабала В.В.

«\_\_\_» \_\_\_\_\_ 2024 р.

\_\_\_\_\_  
(підпис)

**Перевірила:**

доц. кафедри КНІС

к.т.н. Семаньків М.В.

«\_\_\_» \_\_\_\_\_ 2024 р.

\_\_\_\_\_  
(оцінка, підпис)

м. Івано-Франківськ

2024

## Зміст

Зміст	1
Завдання	3
Хід роботи	3
1. Обраний датасет для дослідження	3
2. Попередній аналіз даних:	3
2.1. Завантаження бібліотек та датасету	4
2.2. Кількість рядків та стовпців	4
2.3. Назви стовпців	4
2.4. Типи даних стовпців	4
2.5. Кількість пропущених значень в кожному стовпці	5
2.6. Кількість унікальних даних	6
3. Підготовка даних	8
3.1. Зайва інформація	8
3.2. Зміна типу даних	8
3.3. Видалення анімаційних фільмів без назви	8
3.4. Об'єднані дані	8
3.5. Три стовпці на основі року, місяця та дня з поля 'release_date'	8
4. Дослідження над отриманими даними	10
4.1. Інформація про найстарший анімаційний фільм	10
4.2. Скільки є анімаційних фільмів, що містять в назві слово “story”	10
4.3. Скільки анімаційних фільмів було випущено цього року	10
4.4. Найдовший анімаційний фільм	11
4.5. Скільки анімаційних фільмів вироблено в Україні	11
4.6. Компанія, яка випустила анімаційний фільм з найбільшою популярністю	11
4.7. Середній рейтинг анімаційних фільмів у кожному жанрі	11
4.8. Кількість анімаційних фільмів кожного жанру	12
4.9. Середній бюджет анімаційних фільмів кожної компанії	13
4.10. Середня тривалість анімаційних фільмів для кожного року	13
4.11. Середній рейтинг анімаційних фільмів за мовами	14
4.12. Найвищий бюджет анімаційного фільму для кожного року	14
4.13. Кількість анімаційних фільмів для кожного статусу	15
4.14. Топ-5 компаній за кількістю створених анімаційних фільмів	15
4.15. Анімаційні фільми, що заробили найбільше у кожному жанрі	16
4.16. Середній дохід анімаційних фільмів у кожній країні виробництва	16
4.17. Кількість анімаційних фільмів, випущених у кожному місяці	17

4.18. Кількість анімаційних фільмів випущених кожного дня тижня	18
4.19. Середній рейтинг анімаційних фільмів, що були випущені з 2000 до 2010 року	18
4.20. Найбільший та найменший дохід у кожному році	18
4.21. Кількість анімаційних фільмів з доходом більше 1 мільярда	19
4.22. Кількість анімаційних фільмів з бюджетом менше 100 млн, але доходом більше 500 млн	19
4.23. Кількість мов, якими дубльовані анімаційні фільми	20
4.24. Стовпець profit (прибуток) як різниця між revenue та budget	20
4.25. Стовпець profit_margin (рентабельність), який показує відсоток прибутку від бюджету	20
4.26. Стовпець is_profitable, який показує, чи був анімаційний фільм прибутковим	20
4.27. Зведена таблиця середнього рейтингу по роках випуску та мові	21
4.28. Зведена таблиця сумарного бюджету по місяцю випуску та "adult"-контенту	21
4.29. Зведена таблиця середньої популярності по року та тривалості фільму	22
4.30. Розподіл бюджетів	23
5. Візуалізація отриманих даних	25
5.1. Розподіл 10 найпоширеніших країн-виробників	25
5.2. Розподіл 10 найпоширеніших виробничих компаній	27
5.3. Середня популярність анімаційних фільмів з роками	29
5.4. 15 анімаційних фільмів з найбільшими бюджетами	31
5.5. Кількість анімаційних фільмів випущених по місяцях	33
5.6. Які мови найчастіше використовуються у виробництві анімаційних фільмів?	34
5.7. Залежність прибутку від бюджету	36
5.8. Порівняння кількості прибуткових та неприбуткових анімаційних фільмів	37
5.9. 10 мов з найбільшим середнім прибутком	38
5.10. 5 найдовших за тривалістю українських анімаційних фільмів	40
5.11. Залежність прибутку від тривалості анімаційного фільму	41
5.12. Який анімаційний жанр створюється найбільше	42
5.13. Порівняльний аналіз кількості випущених анімаційних фільмів в період 2001-2010 та 2011-2020 років	44
5.14. Кореляційна матриця основних метрик	46
Письмове обґрунтування результату	48

## Завдання

1. **Обрати датасет для досліджень:** вписати своє прізвище в таблицю з посиланнями
  2. **Вивчити зібрані дані (провести попередній аналіз даних):** типи даних, порожні комірки, унікальні дані, помилкові дані
  3. **Підготовка даних:** відкинути зайву інформацію, змінити типи даних, при потребі заповнити порожні комірки, розділити чи об'єднати дані та оформити остаточний варіант.
  4. **Провести дослідження над отриманими даними.** Створити обчислювальні поля, групування, маски, зведені таблиці, опрацювати статистики та розподіли (всі реалізовані дії підписати)
  5. **Здійснити візуалізацію отриманих даних,** підписати дані, відформатувати. Для окремих результатів досліджень подати окремі графіки чи діаграми.
  6. **Подати коротке письмове обґрунтування** того, який отримано результат дослідження та як ваша візуалізація відображає його.
- Для високих балів доцільно: при візуалізації використати інтерактивність чи анімацію, або застосувати математичний апарат при здійсненні дослідження.

## Хід роботи

1. **Обраний датасет для дослідження**  
[52,000 Animation Movie Details \(2024\)](#)
2. **Попередній аналіз даних:**

## 2.1. Завантаження бібліотек та датасету

```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.animation import FuncAnimation
import matplotlib.animation as animation
import numpy as np

data = pd.read_csv("Animation_Movies.csv")
```

## 2.2. Кількість рядків та стовпців

```
rows, columns = data.shape
print("Number of rows =", rows)
print("Number of columns =", columns)
```

Результат:

```
Number of rows = 51945
Number of columns = 23
```

## 2.3. Назви стовпців

```
print("Names of columns:")
print(data.columns.values)
```

Результат:

```
Names of columns:
['id' 'title' 'vote_average' 'vote_count' 'status'
 'release_date'
 'revenue' 'runtime' 'adult' 'backdrop_path' 'budget'
 'homepage' 'imdb_id'
 'original_language' 'original_title' 'overview' 'popularity'
 'poster_path' 'tagline' 'genres' 'production_companies'
 'production_countries' 'spoken_languages']
```

## 2.4. Типи даних стовпців

```
print("Columns data types", data.dtypes)
```

Результат:

## Columns data types

id	int64
title	object
vote_average	float64
vote_count	int64
status	object
release_date	object
revenue	int64
runtime	int64
adult	bool
backdrop_path	object
budget	int64
homepage	object
imdb_id	object
original_language	object
original_title	object
overview	object
popularity	float64
poster_path	object
tagline	object
genres	object
production_companies	object
production_countries	object
spoken_languages	object

## 2.5. Кількість пропущених значень в кожному стовпці

```
print("Number of Nan values in every column:")
data.isna().sum()
```

Результат:

```
Number of Nan values in every column:
id                0
title             1
vote_average      0
vote_count        0
status            0
release_date     2137
revenue           0
runtime           0
adult             0
backdrop_path    36110
budget            0
homepage         43692
imdb_id          22393
original_language 0
original_title    1
overview         6079
popularity        0
poster_path      14011
tagline          47267
genres            0
production_companies 22547
production_countries 12245
spoken_languages  18127
dtype: int64
```

## 2.6. Кількість унікальних даних

```
print("Number of unique values")
data.nunique()
```

Результат:

```
Number of unique values
id                51945
title             49381
vote_average      2034
vote_count        952
status            6
release_date      16893
revenue           1009
runtime           283
adult             2
backdrop_path     15731
budget            597
homepage          7461
imdb_id           29545
original_language 102
original_title    50318
overview          44907
popularity        5840
poster_path       37819
tagline           4451
genres            2649
production_companies 11196
production_countries 1096
spoken_languages  617
```

На основі проведеного попереднього аналізу даного датасету, можна зробити наступні висновки:

1. **Розмір та структура даних:** Датасет містить 51,945 рядків та 23 стовпці. Це досить великий обсяг інформації, що дозволяє проводити комплексний аналіз та виділяти тенденції в анімаційній індустрії.

## 2. Основні характеристики стовпців:

- Стовпці містять різноманітні типи даних: числові (int64, float64), текстові (object), булеві (bool).
- Дані охоплюють різні аспекти фільмів, такі як назва (title), бюджет (budget), касові збори (revenue), середня оцінка глядачів (vote\_average), тривалість (runtime), інформація про мову (original\_language), жанр (genres), країни виробництва (production\_countries) та інші метадані.

## 3. Пропущені значення:

- Деякі стовпці містять значну кількість пропущених значень. Наприклад, стовпці backdrop\_path (36110 пропусків), homepage (43692), imdb\_id (22393), tagline (47267) мають дуже багато пропущених значень, що вказує на неповноту метаданих.
- Інші стовпці, такі як release\_date та overview, також мають значну кількість пропусків (2137 і 6079 відповідно).

## 4. Унікальні значення:

- Кількість унікальних значень вказує на різноманітність даних. Наприклад, у стовпці title міститься 49,381 унікальна назва, що свідчить про високу унікальність назв фільмів.
- Стовпець vote\_average має 2034 унікальні значення, що дає можливість аналізувати різні рівні оцінок.
- Деякі стовпці мають обмежену кількість унікальних значень, як-от status (6 значень), що вказує на статус випуску фільму (наприклад, "Released", "Post Production" тощо).

## 5. Висновки та можливі проблеми:

- Датасет є досить різноманітним і має великий обсяг даних для аналізу. Однак значна кількість пропущених значень може потребувати обробки (видалення) для коректного аналізу.



### 3. Підготовка даних

#### 3.1. Зайва інформація

Відкинемо інформацію, яка не несе цінності для аналізу та візуалізації

```
data.drop(['backdrop_path', 'poster_path', 'id', 'overview', 'homepage',
          'tagline', 'imdb_id'], axis=1, inplace=True)
```

#### 3.2. Зміна типу даних

Змінимо у стовці 'release\_date' тип даних на datetime

```
data['release_date'] = pd.to_datetime(data['release_date'],
format='%Y-%m-%d')
```

#### 3.3. Видалення анімаційних фільмів без назви

```
data = data.dropna(subset=['title'])
```

#### 3.4. Об'єднані дані

Погрупуємо для майбутніх обчислень та візуалізацій дані за мовою, країною виробництва та компаніями:

```
grouped_by_orig_lang = data.groupby('original_language').size()
grouped__by_prod_countrs = data.groupby('production_countries').size()
grouped_by_prod_comp = data.groupby('production_companies').size()
```

#### 3.5. Три стовпці на основі року, місяця та дня з поля 'release\_date'

```
data['release_year'] = data['release_date'].dt.year
data['release_month'] = data['release_date'].dt.month
data['release_day'] = data['release_date'].dt.day
```

На цьому етапі підготовки даних було виконано кілька важливих кроків, які дозволяють оптимізувати датасет для подальшого аналізу та візуалізації.

#### 1. Видалення зайвої інформації:

- Ми прибрали стовпці, що не несуть значущої аналітичної цінності: backdrop\_path, poster\_path, id, overview, homepage,

tagline та imdb\_id. Це дозволило спростити структуру датасету, залишивши тільки релевантну інформацію для подальших досліджень. Це скорочує обсяг даних та спрощує процес роботи з ними.

**2. Зміна типу даних для release\_date:**

- Зміна типу даних стовпця release\_date на формат datetime дозволяє легко виділяти рік, місяць та день випуску фільму. Це спростить аналіз розподілу фільмів за роками та сезонність випуску анімаційних фільмів.

**3. Видалення фільмів без назви:**

- Видалення записів із відсутніми назвами фільмів гарантує, що всі записи, які ми аналізуватимемо, мають базову ідентифікаційну інформацію. Це допомагає уникнути неповних даних у подальшому аналізі.

**4. Групування даних за мовою, країною та компанією-виробником:**

- Ми створили групи за original\_language, production\_countries та production\_companies. Це дозволить швидко отримувати статистичні дані про кількість фільмів, що випускалися різними мовами, у різних країнах та компаніях, що стане основою для візуалізації та порівняння.

**5. Виділення року, місяця та дня з release\_date:**

- Додавання трьох нових стовпців (release\_year, release\_month та release\_day) дозволяє легко здійснювати аналіз за роком, місяцем та навіть днем випуску. Це відкриває можливості для дослідження сезонності та виявлення тенденцій у випусках фільмів.

Після цього етапу дані стали більш структурованими та зручними для аналізу. Видалення зайвої інформації та перетворення дат відкриває більше можливостей для точнішого і швидшого аналізу анімаційних фільмів за різними аспектами, зокрема, за роком, країною, мовою та компанією-виробником.

## 4. Дослідження над отриманими даними

### 4.1. Інформація про найстарший анімаційний фільм

```
data.loc[data["release_date"].argmin()]
```

Результат:

title	The Musician Monkey
vote_average	6.048
vote_count	21
status	Released
release_date	1878-05-07 00:00:00
revenue	0
runtime	1
adult	False
budget	153
original_language	xx
original_title	Le Singe Musicien
popularity	1.031
genres	Animation
production_companies	NaN
production_countries	France
spoken_languages	No Language

### 4.2. Скільки є анімаційних фільмів, що містять в назві слово “story”

```
data[data['title'].str.contains('story', case=False)].shape[0]
```

Результат:

477

### 4.3. Скільки анімаційних фільмів було випущено цього року

```
data[(data["release_date"].dt.year == 2024) & (data["status"] == "Released")].shape[0]
```

Результат:

52

#### 4.4. Найдовший анімаційний фільм

```
data.loc[data["runtime"].argmax()]["title"]
```

Результат:

Giant 600 Cartoon Collection

Цей фільм має найбільшу тривалість у нашому датасеті. Це, ймовірно, збірка анімацій, оскільки така тривалість є надзвичайно великою для одного фільму.

#### 4.5. Скільки анімаційних фільмів вироблено в Україні

```
data[(data["production_countries"] == "Ukraine")].shape[0]
```

Результат:

273

Кількість анімаційних фільмів українського виробництва становить приблизно 0.55% від загальної кількості фільмів в датасеті.

#### 4.6. Компанія, яка випустила анімаційний фільм з найбільшою популярністю

```
data.loc[data["popularity"].argmax()]["production_companies"]
```

Результат:

Walt Disney Pictures, Pixar

Найпопулярніший анімаційний фільм у датасеті ([Elemental](#)) випущений компаніями Walt Disney Pictures та Pixar, що підтверджує їхній вплив на анімаційну індустрію та популярність їхніх фільмів.

#### 4.7. Середній рейтинг анімаційних фільмів у кожному жанрі

```
data['genres'] = data['genres'].str.split(', ')
data.explode('genres').groupby('genres')['vote_average'].mean().sort_values(ascending=False)
```

Результат:

Action	4.901401	Mystery	4.724864
TV Movie	4.803455	Adventure	4.646531
		Crime	4.621981

Science Fiction	4.333737	Western	3.942422
Fantasy	4.272016	Drama	3.771976
Thriller	4.236690	Music	3.203602
Comedy	4.226729	History	3.078425
Family	4.221120	Horror	3.060209
		Animation	2.597305
Romance	4.017737	Documentary	1.840423
War	4.013777		

Найвищий середній рейтинг мають жанри Action, TV Movie, Mystery; найнижчий — Animation та Documentary.

Жанр Animation має нижчий середній рейтинг порівняно з іншими жанрами. Це може бути зумовлено тим, що до категорії анімації входять фільми різної якості й призначені для різних вікових категорій.

#### 4.8. Кількість анімаційних фільмів кожного жанру

```
data.explode('genres').groupby('genres').size().sort_values(ascending=False)
```

Результат

Animation	51953	Horror	1393
Comedy	7879	Romance	866
Family	7480	TV Movie	679
Fantasy	3896	Mystery	543
Adventure	3539	History	499
Drama	2849	War	368
Science Fiction	2540	Crime	320
Action	2428	Thriller	306
Documentary	1897	Western	161
Music	1697		

Найбільше фільмів у жанрах Animation, Comedy, та Family; найменше — у Western та Thriller.

Найбільша кількість фільмів у жанрі Animation свідчить про загальну популярність анімаційного жанру. Comedy та Family теж популярні, що вказує на переваги аудиторії в анімаційних фільмах для сімейного перегляду та розваг.

#### 4.9. Середній бюджет анімаційних фільмів кожної компанії

```
data['production_companies'] =
data['production_companies'].str.split(',')
data.explode('production_companies').groupby('production_companies')
['budget'].mean().sort_values(ascending=False)
```

Результат:

Golden Mean	165000000.0
Jerry Bruckheimer Films	150000000.0
Proximity Media	150000000.0
SpringHill Entertainment	150000000.0
Whamaphram Productions	150000000.0
...	
Gruppo Memmo La Roca	0.0
Guang Dong Animation	0.0
GuangZhou Tongjiatong Education Consulting	0.0
Guangdong Original Power Culture Communication	0.0
한국독립애니메이션협회	0.0

Найвищий середній бюджет мають компанії, такі як Golden Mean, Jerry Bruckheimer Films, Proximity Media, та Whamaphram Productions, які виділяли значні кошти на виробництво анімаційних фільмів.

Це може свідчити про амбіційні проекти або про велику увагу до якості й масштабу анімаційних фільмів з боку цих компаній.

#### 4.10. Середня тривалість анімаційних фільмів для кожного року

```
data.groupby('release_year')['runtime'].mean()
```

Результат:

1878.0	1.000000
1885.0	1.000000
1892.0	9.666667
1894.0	7.000000
1896.0	1.000000
...	
2022.0	17.279484
2023.0	20.047719
2024.0	21.310680
2025.0	18.200000
2026.0	37.500000

Тривалість анімаційних фільмів зростає з часом. Починаючи з мінімальної тривалості у 1800-х роках, вона поступово збільшується до 20–30 хвилин у сучасний період. Це пов'язано з еволюцією технологій та розвитком індустрії, що дозволяє створювати довші й складніші анімаційні сюжети.

#### 4.11. Середній рейтинг анімаційних фільмів за мовами

```
data.groupby('original_language')['vote_average'].mean().sort_values(ascending=False)
```

Результат:

te	9.0	gn	0.0
as	8.0	my	0.0
oc	8.0	qu	0.0
si	7.6	jv	0.0
sa	7.5	lb	0.0

Найвищий середній рейтинг мають мови, такі як Телугу, Ассамська, Окситанська, Сингальська, та Санскрит, тоді як деякі мови мають нульовий рейтинг.

Високі рейтинги фільмів цими мовами найімовірніше свідчать про те, що декілька людей поставили максимальний бал цим анім. фільмам. Мови з нульовим рейтингом можуть вказувати на те, що на них було знято мало фільмів або вони не отримали достатньої уваги від глядачів.

#### 4.12. Найвищий бюджет анімаційного фільму для кожного року

```
data.groupby('release_year')['budget'].max()
```

Результат:

1878.0	153
1885.0	0
1892.0	0
1894.0	0
1896.0	0
...	
2022.0	200000000
2023.0	200000000
2024.0	40000000

```
2025.0      30000000
2026.0           0
```

Максимальний бюджет анімаційного фільму на рік коливається, досягаючи 200 мільйонів доларів у 2022–2023 роках.

Це відображає загальну тенденцію зростання бюджетів з часом, що свідчить про збільшення інвестицій у анімаційні фільми та підвищення вимог до їх якості.

#### 4.13. Кількість анімаційних фільмів для кожного статусу

```
data['status'].value_counts()
```

Результат:

```
Released      51133
In Production   381
Planned        305
Post Production  97
Canceled       17
Rumored        11
```

Більшість фільмів мають статус Released (випущені) — 51133. Також є фільми, які знаходяться на різних етапах виробництва або були скасовані. Така велика кількість випущених фільмів свідчить про стабільне виробництво, але також видно, що деякі фільми зупинені на різних етапах з різних причин.

#### 4.14. Топ-5 компаній за кількістю створених анімаційних фільмів

```
data.explode('production_companies').groupby('production_companies').size().nlargest(5)
```

Результат:

```
ONF | NFB      847
Walt Disney Productions  716
Soyuzmultfilm    681
Walter Lantz Productions  560
Warner Bros. Pictures    533
```



#### 4.15. Анімаційні фільми, що заробили найбільше у кожному жанрі

```
data.explode('genres').sort_values('revenue',
ascending=False).drop_duplicates('genres')[['genres',
'title', 'revenue']]
```

Результат:

	genres	title	revenue
40	Comedy	Frozen II	1450026933
40	Adventure	Frozen II	1450026933
40	Animation	Frozen II	1450026933
40	Family	Frozen II	1450026933
40	Fantasy	Frozen II	1450026933
23	Action	Incredibles 2	1242805359
0	Drama	Inside Out	857611174
3	Music	Coco	800526015
4	Science Fiction	WALL-E	521311860
28	Romance	Aladdin	504050219
156	Thriller	Demon Slayer -Kimetsu no Yaiba- The Movie: Mug...	503063688
104	Mystery	The Adventures of Tintin	373993951
100	Crime	Who Framed Roger Rabbit	329803958
81	Western	Rango	245724603
323	Horror	The Addams Family 2	119815153
183	War	The Wind Rises	117932401
205	History	Loving Vincent	42187665
397	Documentary	Waltz with Bashir	11179372
41152	TV Movie	Balala The Fairies	7600000

#### 4.16. Середній дохід анімаційних фільмів у кожній країні виробництва

```
data['production_countries'] =
data['production_countries'].str.split(', ')
data.explode('production_countries').groupby('production_coun
tries')['revenue'].mean().sort_values(ascending=False)
```

Результат:

```
production_countries
New Zealand          7.804099e+06
United States of America  6.066280e+06
Australia            4.932477e+06
China                4.521717e+06
Denmark              3.603571e+06
...
Holy See             0.000000e+00
Haiti                0.000000e+00
Guyana               0.000000e+00
Guatemala            0.000000e+00
Zimbabwe             0.000000e+00
Name: revenue, Length: 154, dtype: float64
```

Найвищий середній дохід анімаційних фільмів припадає на Нову Зеландію, США, Австралію, Китай та Данію.

Це може свідчити про високий рівень розвитку анімаційної індустрії в цих країнах, особливо в США, які є глобальним лідером у кіновиробництві.

#### 4.17. Кількість анімаційних фільмів, випущених у кожному місяці

```
data['release_month'].value_counts().sort_index()
```

Результат:

```
release_month
1.0      11959
2.0       2837
3.0       3557
4.0       2958
5.0       3139
6.0       4063
7.0       3101
8.0       2848
9.0       3783
10.0      4250
11.0      3611
12.0      3701
Name: count, dtype: int64
```

Найбільше фільмів випущено в січні, жовтні та червні, тоді як лютий і серпень мають менші показники. Сезонність у випуску анімаційних

фільмів може бути пов'язана з маркетинговими стратегіями, орієнтованими на канікули та святкові періоди.

#### 4.18. Кількість анімаційних фільмів випущених кожного дня тижня

```
data['release_day_of_week'] = data['release_date'].dt.day_name()
data['release_day_of_week'].value_counts()
```

Результат:

```
release_day_of_week
Friday      9494
Saturday    8134
Thursday    7673
Tuesday     6760
Wednesday   6538
Monday      5617
Sunday      5591
Name: count, dtype: int64
```

Найчастіше фільми випускалися у п'ятницю та суботу. Популярність цих днів обумовлена тим, щоб охопити аудиторію на вихідних.

#### 4.19. Середній рейтинг анімаційних фільмів, що були випущені з 2000 до 2010 року

```
data[(data['release_year'] > 2000) & (data['release_year'] < 2010)]['vote_average'].mean()
```

Результат:

```
3.1388660234103134
```

Середній рейтинг становить приблизно 3.14. Це показує стабільний інтерес до анімацій у цей період, але середній рейтинг може свідчити про різний рівень якості випущених фільмів.

#### 4.20. Найбільший та найменший дохід у кожному році

```
data.groupby('release_year')['revenue'].agg(['max', 'min'])
```

Результат:

```
      max    min
release_year
```

```

1878.0    0    0
1885.0    0    0
1892.0    0    0
1894.0    0    0
1896.0    0    0
...    ...    ...
2022.0    939628210  0
2023.0    1355725263  0
2024.0     11    0
2025.0     0    0
2026.0     0    0
130 rows x 2 columns

```

Максимальні доходи зросли до 1.35 мільярда у 2023 році, тоді як мінімальні залишаються на рівні 0 для багатьох років. Зростання максимального доходу показує, що індустрія анімаційних фільмів стала більш прибутковою з часом, але не всі проекти були комерційно успішними.

#### 4.21. Кількість анімаційних фільмів з доходом більше 1 мільярда

```
data[data['revenue'] > 1e9].shape[0]
```

Результат:

```
10
```

Лише 10 фільмів досягли доходу понад 1 мільярд. Це свідчить про рідкість настільки високих касових зборів, що характерно для найбільших хітів.

#### 4.22. Кількість анімаційних фільмів з бюджетом менше 100 млн, але доходом більше 500 млн

```
data[(data['budget'] < 1e8) & (data['revenue'] > 5e8)][['title',
'budget', 'revenue']].shape[0]
```

Результат:

```
20
```

20 фільмів мали відносно невеликий бюджет, але принесли значні прибутки. Ці фільми є прикладами високої ефективності інвестицій,

показуючи, що великий успіх можливий і з помірними витратами на виробництво.

#### 4.23. Кількість мов, якими дубльовані анімаційні фільми

```
data['spoken_languages'] = data['spoken_languages'].str.split(',')
data.explode('spoken_languages')['spoken_languages'].nunique()
```

Результат:

115

Фільми дубльовано на 115 мовах. Це свідчить про глобальне охоплення анімаційної продукції, яке орієнтоване на широку аудиторію в усьому світі.

#### 4.24. Стовпець profit (прибуток) як різниця між revenue та budget

```
data['profit'] = data['revenue'] - data['budget']
```

Цей стовпець дозволяє оцінити, наскільки прибутковим був кожен анімаційний фільм у абсолютних числах. Високі показники прибутку вказують на комерційно успішні фільми, тоді як від'ємні значення свідчать про збитки.

#### 4.25. Стовпець profit\_margin (рентабельність), який показує відсоток прибутку від бюджету

```
data['profit_margin'] = data['profit'] / data['budget'] * 100
```

Це значення показує ефективність використання бюджету кожного фільму. Високий відсоток рентабельності означає, що фільм отримав значний прибуток порівняно з вкладеннями.

#### 4.26. Стовпець is\_profitable, який показує, чи був анімаційний фільм прибутковим

```
data['is_profitable'] = data['profit'] > 0
```

Цей бінарний стовпець дозволяє швидко визначити, чи досяг фільм прибутковості. Він може використовуватися для розрахунку загальної частки прибуткових анімаційних фільмів у вибірці, що показує успішність галузі загалом.

#### 4.27. Зведена таблиця середнього рейтингу по роках випуску та мові

```
last_10_years = data[(data['release_year'] >= (data['release_year'].max()
- 12)) & (data['release_year'] <= 2024.0)]
languages = ['en', 'ja', 'fr', 'es', 'de']
filtered_data =
last_10_years[last_10_years['original_language'].isin(languages)]
pivot_table = pd.pivot_table(filtered_data, values='vote_average',
index='release_year', columns='original_language', aggfunc='mean')
```

Результат:

original_language	de	en	es	fr	ja
release_year					
2014.0	2.919759	3.079226	3.252043	3.530239	3.991994
2015.0	3.377419	3.208271	3.092945	3.812046	4.253966
2016.0	2.931943	2.874459	2.494667	3.528747	3.817347
2017.0	1.584214	2.482988	3.311044	3.325952	3.546777
2018.0	3.259467	2.360591	3.782627	3.406136	3.337025
2019.0	2.386933	2.065012	1.988913	3.428671	3.772449
2020.0	1.425000	1.843741	1.728917	2.752920	2.887529
2021.0	0.232653	1.629947	1.696097	2.002865	3.176642
2022.0	1.042260	1.573144	1.537932	1.907582	4.142123
2023.0	0.400000	1.096155	1.190294	1.365475	2.952717
2024.0	0.000000	0.163934	0.000000	0.000000	0.000000

Ця таблиця показує, як змінювався середній рейтинг фільмів різними мовами з часом. Наприклад, можна побачити, які мови демонструють стабільно високий чи низький рейтинг у певні роки, що може вказувати на якість або популярність контенту для цієї мови.

#### 4.28. Зведена таблиця сумарного бюджету по місяцю випуску та "adult"-контенту

```
pd.pivot_table(data, values='budget', index='release_month',
columns='adult', aggfunc='mean')
```

Результат:

	adult	False	True
release_month			
1.0	6.294584e+04	232.558140	
2.0	5.414187e+05	100.000000	
3.0	9.000243e+05	13.157895	
4.0	3.410253e+05	0.000000	
5.0	6.479720e+05	0.000000	
6.0	1.513693e+06	1346.437500	
7.0	9.343374e+05	0.000000	
8.0	3.703564e+05	0.000000	
9.0	4.094829e+05	0.000000	
10.0	6.288300e+05	23.076923	
11.0	1.295871e+06	0.000000	
12.0	5.837104e+05	7285.714286	

Ця зведена таблиця вказує на середній бюджет анімаційних фільмів, розділених за місяцями випуску та категорією "дорослий контент". Видно, що фільми для широкої аудиторії мають значно вищий середній бюджет порівняно з дорослими, що свідчить про орієнтацію індустрії на сімейний контент. Також бюджет у червні та листопаді для фільмів без маркування "adult" є досить високим, що може вказувати на сезонність випусків більш дорогих проектів.

#### 4.29. Зведена таблиця середньої популярності по року та тривалості фільму

```
pd.pivot_table(last_10_years, values='popularity',
index='release_year', columns=pd.cut(data['runtime'], bins=[0, 60,
90, 120, 150, 200]), aggfunc='mean')
```

Результат:

runtime	(0, 60]	(60, 90]	(90, 120]	(120, 150]	(150, 200]
release_year					
2014.0	1.032255	5.202354	11.430636	1.787700	3.461000
2015.0	0.990034	4.866437	10.972247	2.024500	0.600000
2016.0	0.873776	5.334673	11.474959	9.832750	0.622500
2017.0	0.927404	4.927293	11.853128	2.912250	0.600000
2018.0	0.810610	4.179144	9.412603	2.276500	0.600000
2019.0	0.940079	4.018521	11.316333	1.581500	7.859400
2020.0	0.839688	5.304788	12.341379	8.006556	6.428500
2021.0	0.973091	7.627429	17.610656	13.212444	3.777875
2022.0	1.075489	9.916000	24.316989	25.961857	2.077667
2023.0	2.684955	11.605680	47.568062	37.784267	6.110250
2024.0	0.285061	1.896300	2.173500	3.225667	0.000000

Ця таблиця показує середню популярність анімаційних фільмів за певні роки та категорії тривалості. З неї можна зробити висновок, що популярність фільмів різної тривалості змінюється з роками, і фільми середньої тривалості (від 90 до 150 хвилин) мають стабільно високий рівень популярності.

#### 4.30. Розподіл бюджетів

```
data['budget'].describe()
```

Результат:

```
count    5.194400e+04
mean     5.759624e+05
std      7.606701e+06
min      0.000000e+00
25%      0.000000e+00
50%      0.000000e+00
75%      0.000000e+00
max      2.600000e+08
```



Аналіз розподілу бюджетів анімаційних фільмів показує наступне:

- Кількість фільмів із зазначеним бюджетом: Є дані про бюджети для 51,944 анімаційних фільмів.
- Середній бюджет: 575,962, що свідчить про значні коливання між фільмами низького та високого бюджету.
- Стандартне відхилення: 7,606,701 — високе значення, яке показує велику варіативність у бюджетах, від малобюджетних до значно фінансованих проєктів.
- Мінімальний бюджет: 0, що може включати незалежні чи аматорські проєкти.
- Медіана та квартилі: Для 75% фільмів бюджет становить 0, що вказує на значну кількість проєктів без офіційного бюджету або з надзвичайно низьким фінансуванням.
- Максимальний бюджет: 260 мільйонів, ймовірно, для великих студійних проєктів.

Висновок: Сильний перекис у розподілі бюджетів свідчить про наявність багатьох малобюджетних фільмів поруч із невеликою кількістю високобюджетних.

## 5. Візуалізація отриманих даних

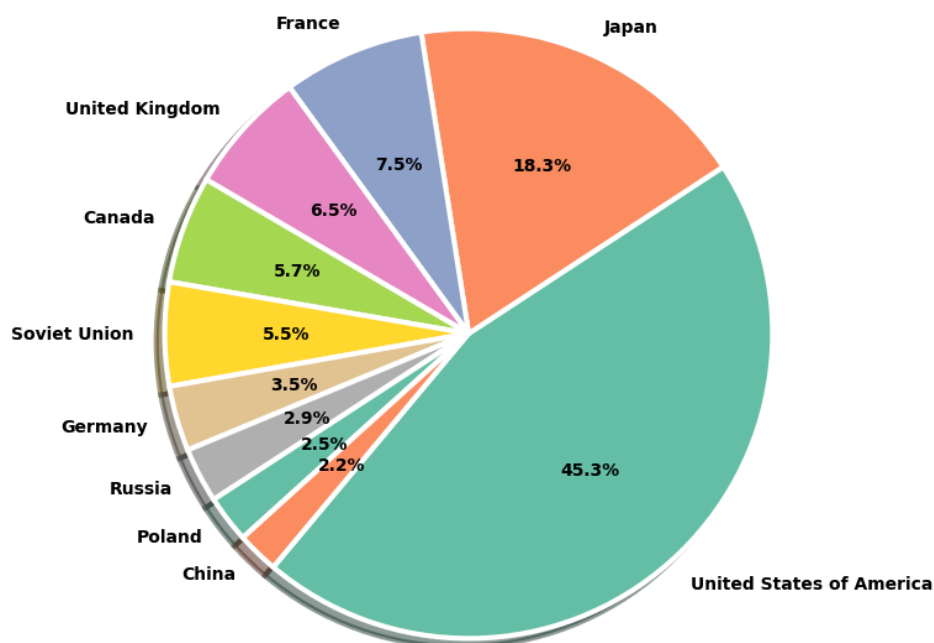
### 5.1. Розподіл 10 найпоширеніших країн-виробників

```
fig, ax = plt.subplots(figsize=(10, 8))
def update(num):
    ax.clear()
    ax.pie(
        grouped_by_prod_countrs,
        labels=grouped_by_prod_countrs.index,
        autopct="%1.1f%%",
        shadow=True,
        textprops={"fontsize": 10, "weight": "bold"},
        colors=sns.color_palette("Set2"),
        startangle=90 + num * 4,
        wedgeprops={"linewidth": 3, "edgecolor": "white"},
    )
    ax.set_title(
        "Distribution of the 10 most common production countries",
        fontsize=18,
        fontweight="bold",
    )
ani = FuncAnimation(fig, update, frames=np.arange(0, 36),
interval=400)
ani.save("5.1.gif", writer="imagemagick", fps=5)

plt.show()
```

[Анімований результат:](#)

#### Distribution of the 10 most common production countries



Графік показує розподіл десяти найбільш поширених країн-виробників анімаційних фільмів, виражений у відсотках. Основні висновки:

1. **США** домінує серед країн-виробників, відповідаючи за **45.3%** усіх анімаційних фільмів. Це пояснюється потужною індустрією анімації в США, де працюють такі великі студії, як Disney, Pixar і DreamWorks, що мають значний вплив на ринок анімації.
2. **Японія** займає друге місце з **18.3%**. Японська анімація або аніме, користується широкою популярністю як в країні, так і за її межами, завдяки таким студіям, як Studio Ghibli, Toei Animation та інші.
3. **Франція** займає третю позицію з **7.5%**. Французька анімаційна індустрія також розвинена та добре підтримується на державному рівні. Студії, як-от Xilam та Folimage, роблять свій внесок у світ анімації.
4. **Велика Британія** (6.5%), **Канада** (5.7%), і **Радянський Союз** (5.5%) також мають значний внесок у виробництво анімаційних фільмів. Велика Британія та Канада відомі як основні виробники анімації завдяки підтримці незалежних студій, а Радянський Союз мав традицію державної анімації.
5. Інші країни, такі як **Німеччина** (3.5%), **росія** (2.9%), **Польща** (2.5%), та **Китай** (2.2%), мають менші частки на ринку. Проте вони також зробили свій внесок у світову анімацію через власні культурні та художні підходи.

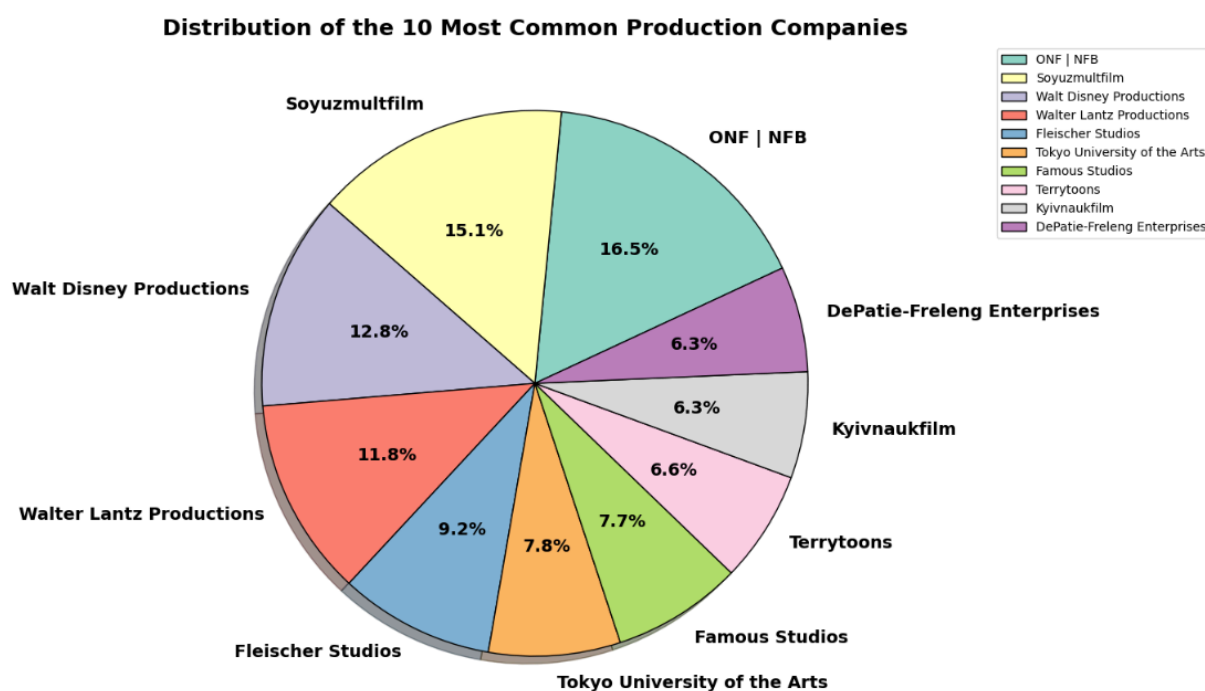
Отже, ринок анімації є досить концентрованим у кількох основних країнах, таких як США, Японія та Франція. Ці країни задають тон розвитку анімації та експортують свою продукцію по всьому світу, що сприяє популяризації їхньої культури та впливу в інших країнах.

## 5.2. Розподіл 10 найпоширеніших виробничих компаній

```
fig, ax = plt.subplots(figsize=(18, 10))
explode = [0.2] * len(grouped_by_prod_comp)
cmap = plt.get_cmap("Set3")
num_colors = len(grouped_by_prod_comp.index)
cmap(np.arange(num_colors) * 2)
def animate(i):
    ax.clear()
    counts = grouped_by_prod_comp
    labels = grouped_by_prod_comp.index
    explode[i % len(explode)] = 0
    colors[i % num_colors] = cmap(i % num_colors)
    ax.pie(counts, labels=labels, explode=explode, autopct="%1.1f%%",
           wedgeprops={ "edgecolor": "k", "linewidth": 1, "linestyle":
"solid", "antialiased": True, }, textprops={"fontsize": 14, "weight":
"bold"}, shadow=True, startangle=25, colors=colors,)
    explode[i % len(explode)] = 0
    ax.legend(bbox_to_anchor=(1.5, 1), loc="upper right")
    ax.set_title("Distribution of the 10 Most Common Production
Companies", fontsize=18, fontweight="bold",)

ani = FuncAnimation(fig, animate, frames=len(explode), interval=500,
repeat=False)
ani.save("5.2.gif", writer="imagemagick", fps=5)
```

### Анімований результат:



Графік показує розподіл десяти найпоширеніших анімаційних виробничих компаній за їх часткою у загальному обсязі виробництва анімаційних фільмів. Основні висновки з аналізу:

1. **ONF | NFB (Національна кінорада Канади)** займає найбільшу частку з **16.5%**. Це свідчить про активну участь Канади у виробництві анімаційних фільмів, де Національна кінорада підтримує багато освітніх, культурних і комерційних проектів.
2. **Soyuzmultfilm** – одна з найбільших анімаційних студій Радянського Союзу та Росії – займає **15.1%**. Вона є відомою студією, що створила багато класичних анімаційних фільмів у Радянському Союзі, які мали великий вплив у межах країни.
3. **Walt Disney Productions** має **12.8%** частки. Це відображає значний внесок компанії Disney у світ анімації, оскільки вона є однією з найбільших і найвпливовіших студій у світі, що постійно випускає популярні та культові анімаційні фільми.
4. **Walter Lantz Productions** – виробник мультфільмів про [Вуді Вудпекера](#) – займає **11.8%**. Хоча ця студія не така велика, як Disney, вона все ж має важливе місце в історії американської анімації.
5. **Fleischer Studios** з **9.2%** відома своєю класичною анімацією, як-от [Бетті Буп](#) і [Попай](#). Студія мала значний вплив на ранній етап розвитку анімаційної індустрії США.
6. **Tokyo University of the Arts** (7.8%) та **Famous Studios** (7.7%) також зробили вагомий внесок у анімацію. Tokyo University of the Arts випускає молодих аніматорів, які згодом працюють у великих студіях, а Famous Studios продовжувала традиції Fleischer Studios, створюючи класичні американські мультфільми.
7. **Terrytoons**, **Kyivnaukfilm** і **DePatie-Freleng Enterprises** займають приблизно по **6.3-6.6%** кожна. Kyivnaukfilm відома науково-популярною та навчальною анімацією, Terrytoons – американською анімацією, що створила таких персонажів, як [Mighty Mouse](#), а DePatie-Freleng – анімаційними проектами для телебачення, зокрема [Рожева Пантера](#).

На графіку видно значну різноманітність у виробництві анімації різними студіями з різних країн. ONF | NFB, Soyuzmultfilm і Walt Disney Productions є основними гравцями, що мають суттєву частку ринку.

### 5.3. Середня популярність анімаційних фільмів з роками

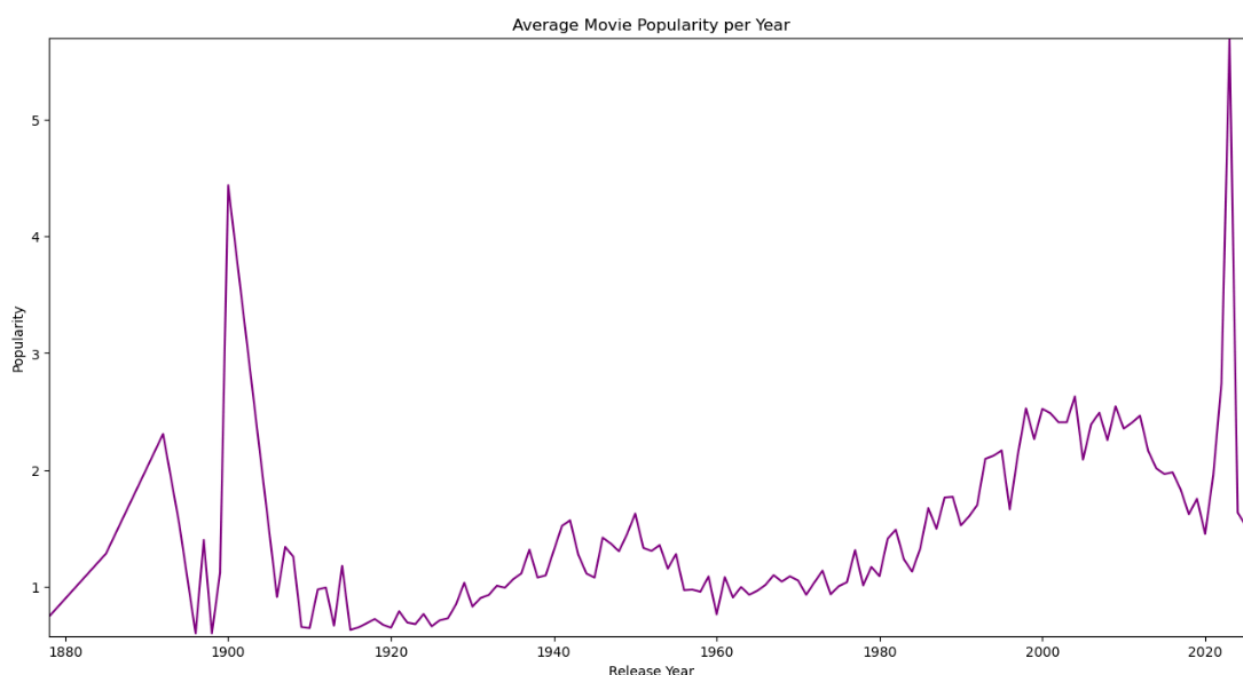
```

avg_popularity_per_year =
data.groupby('release_year')['popularity'].mean()
fig, ax = plt.subplots(figsize=(16, 8))
ax.set_title('Average Movie Popularity per Year')
ax.set_xlabel('Release Year')
ax.set_ylabel('Popularity')
line, = ax.plot([], [], color='purple')

def init():
    ax.set_xlim(avg_popularity_per_year.index.min(),
avg_popularity_per_year.index.max())
    ax.set_ylim(avg_popularity_per_year.min(),
avg_popularity_per_year.max())
    line.set_data([], [])
    return line,
def animate(i):
    x = avg_popularity_per_year.index[:i+1]
    y = avg_popularity_per_year.values[:i+1]
    line.set_data(x, y)
    return line,
ani = FuncAnimation(fig, animate, frames=len(avg_popularity_per_year),
init_func=init, blit=True)
ani.save('5.3.gif', writer='imagemagick')
plt.show()

```

[Анімований результат:](#)



Графік показує зміну середньої популярності анімаційних фільмів з 1880-х років до 2020-х. Основні тенденції:

1. **Ранній період (до 1920-х років):** починаючи з 1880-х років, популярність анімаційних фільмів була відносно низькою, з кількома піками, особливо в період до 1900 року. Ці коливання можуть відображати зростання інтересу до нових кінематографічних технологій того часу, а також експериментальну природу перших анімацій.
2. **Стабільний ріст (після 1920-х років):** після 1920-х років популярність фільмів залишається на досить стабільному рівні з незначними коливаннями. Це може бути пов'язано з тим, що анімація ще не досягла значного рівня комерційного розвитку та залишалася нішевою сферою кінематографу.
3. **Підйом популярності (1960–2000):** з 1960-х років популярність анімаційних фільмів починає зростати. Це може бути пов'язано з появою відомих студій, таких як Walt Disney, і зростанням попиту на анімаційні фільми. Більше глядачів зацікавлені у мультфільмах, а студії розширюють виробництво, зокрема на телевізійний ринок.
4. **Різкий зріст (2000–2020):** починаючи з 2000-х років, спостерігається стрімке зростання популярності анімаційних фільмів. Це пов'язано з розвитком комп'ютерної анімації, успіхом таких компаній, як Pixar, DreamWorks, та доступністю фільмів через цифрові платформи. Новітні технології дозволили створювати більш привабливі та якісні анімаційні фільми, що привертають широку аудиторію.
5. **Пік у 2020-х:** на графіку видно значний стрибок популярності в 2020 році, ймовірно, через стрімінгові сервіси, які стали дуже популярними під час пандемії COVID-19. Попит на контент зріс, і анімація, як універсальний жанр для різних вікових груп, стала ще більш популярною.

Отже, можемо зробити висновки, що популярність анімаційних фільмів з роками значно зросла, особливо в останні десятиліття завдяки технічним досягненням та зростанню доступу до контенту через стрімінгові платформи. Тенденція до зростання свідчить про те, що анімація продовжує залишатися важливим і популярним жанром кіноіндустрії.

## 5.4. 15 анімаційних фільмів з найбільшими бюджетами

```
highest_budget_movies = data[['title', 'budget']].sort_values(by='budget',
ascending=False).head(15)
fig, ax = plt.subplots(figsize=(12, 6))
ax.set_title("Top 15 Highest-Budget Movies")
ax.set_xlabel('Budget [mld]')
ax.set_ylabel("Movie Title")
sns.set(style="whitegrid")

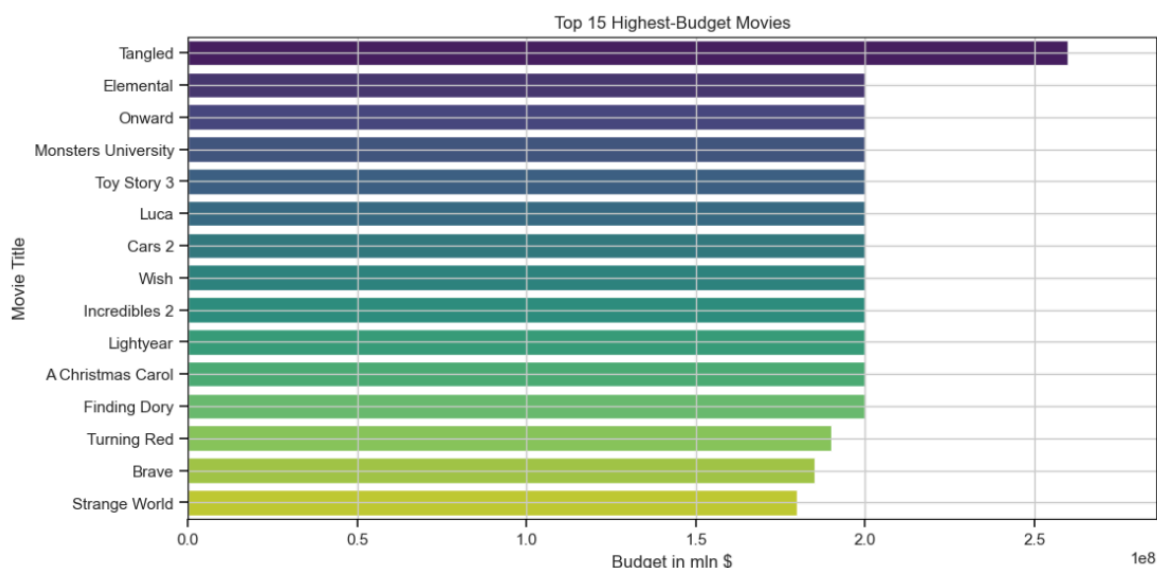
def init():
    ax.clear()
    sns.barplot(data=highest_budget_movies, y='title', x='budget',
orient='h', ax=ax)
    ax.set_xlim(0, highest_budget_movies['budget'].max() * 1.1)

def animate(i):
    ax.clear()
    current_data = highest_budget_movies.head(i + 1)
    sns.barplot(data=current_data, y='title', x='budget', orient='h',
ax=ax, palette="viridis")
    ax.set_xlim(0, highest_budget_movies['budget'].max() * 1.1)
    ax.set_title("Top 15 Highest-Budget Movies")
    ax.set_xlabel('Budget in mln $')
    ax.set_ylabel("Movie Title")

anim = FuncAnimation(fig, animate, frames=len(highest_budget_movies),
init_func=init, interval=500, repeat=False)

anim.save('5.4.gif', writer='imagemagick')
plt.show()
```

Анімований результат:





Наведений графік показує 15 анімаційних фільмів із найбільшими бюджетами. Візуалізація демонструє назви фільмів на вертикальній осі та їхні бюджети на горизонтальній осі, виражені в мільйонах доларів (млн). Основні тенденції:

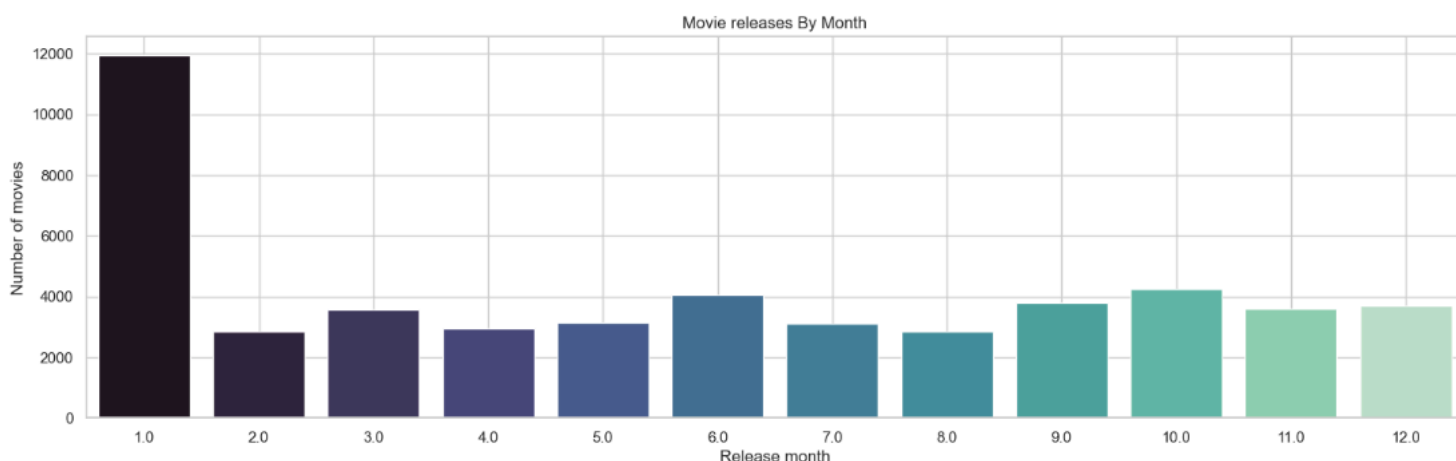
1. **Лідер списку:** фільм "[Tangled](#)" має найбільший бюджет серед анімаційних стрічок у цьому списку. Його бюджет значно перевищує бюджети інших фільмів і наближається до 260 мільйонів доларів.
2. **Високі бюджети:** усі фільми, представлені на графіку, мають бюджети вище 150 мільйонів доларів. Це свідчить про значні інвестиції, необхідні для створення високоякісних анімаційних фільмів.
3. **Різноманітність жанрів і тем:** представлені фільми належать до різних жанрів та мають різні стилі анімації (наприклад, "[Monsters University](#)", "[Toy Story 3](#)", "[Cars 2](#)" та "[Incredibles 2](#)"), що показує, що високий бюджет не обмежується лише певним типом анімації.
4. **Схожі бюджети:** фільми після "Tangled" мають відносно схожі бюджети, які коливаються в межах 150-200 мільйонів доларів, що вказує на певний стандартний рівень витрат для студій, таких як Disney та Pixar, на створення своїх проектів.
5. **Новіші та старіші релізи:** на графіку є як новіші фільми ("[Elemental](#)", "[Wish](#)"), так і старіші ("Toy Story 3", "[A Christmas Carol](#)"). Це свідчить про те, що індустрія анімації постійно розвивається і готова вкладати значні кошти в нові проекти.

Отже, графік демонструє значні витрати на створення анімаційних фільмів, які студії вкладають, щоб забезпечити високу якість і конкурентоспроможність на ринку.

### 5.5. Кількість анімаційних фільмів випущених по місяцях

```
release =
data["release_month"].value_counts().sort_values(ascending=False)
plt.figure(figsize=(15,5))
sns.barplot(x=release.index, y=release.values, palette="mako")
plt.title("Movie releases By Month")
plt.grid(True)
plt.show()
```

Результат:



Дана візуалізація відображає кількість анімаційних фільмів, випущених по місяцях протягом року. Основні тенденції:

1. **Найбільша кількість релізів у січні:** січень (позначений як 1.0 на осі X) виділяється як місяць із найбільшою кількістю випущених анімаційних фільмів — майже 12 000. Це може бути пов'язано з тим, що початок року часто використовується для прем'єр, спрямованих на привернення уваги нових аудиторій або для новорічних святкових періодів.
2. **Інші місяці рівномірні:** в усі інші місяці кількість випущених фільмів коливається в межах приблизно від 2 000 до 4 000. Це вказує на рівномірний розподіл релізів протягом року, без чітко виражених піків, окрім січня.
3. **Помірні підвищення в жовтні та червні:** спостерігається невелике збільшення кількості релізів у жовтні та червні, що може бути пов'язано з літнім періодом відпусток та підготовкою до святкових прем'єр восени.

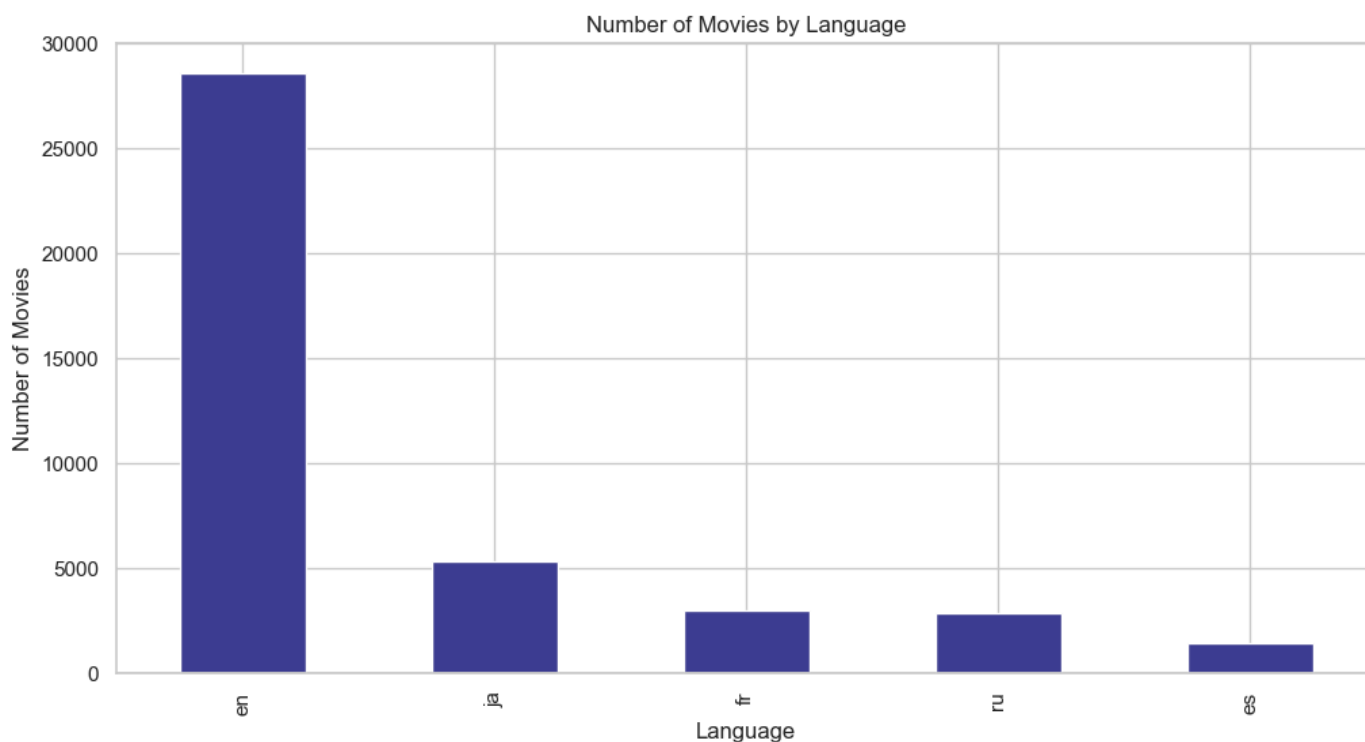
4. **Поступове зростання восени:** також можемо помітити незначне збільшення кількості випусків у вересні та жовтні. Це може бути пов'язано з підготовкою до святкових релізів та прагненням охопити глядацьку аудиторію перед кінцем року.
5. **Зниження в грудні:** у грудні кількість випущених фільмів трохи знижується порівняно з осінніми місяцями, що може бути пов'язано з тим, що більші релізи плануються на початок нового року або на інші свята.

Загалом, графік показує, що січень є найпопулярнішим місяцем для релізу анімаційних фільмів, тоді як інші місяці характеризуються рівномірним розподілом релізів без чітко виражених піків.

### 5.6. Які мови найчастіше використовуються у виробництві анімаційних фільмів?

```
plt.figure(figsize=(12, 6))
data['original_language'].value_counts().head(5).plot(kind='bar',
color='#404094')
plt.title('Number of Movies by Language')
plt.xlabel('Language')plt.ylabel('Number of Movies')
plt.show()
```

Результат:



На даній візуалізації показано кількість анімаційних фільмів, випущених різними мовами. На горизонтальній осі представлено мови, а на вертикальній – кількість фільмів, випущених на цих мовах.

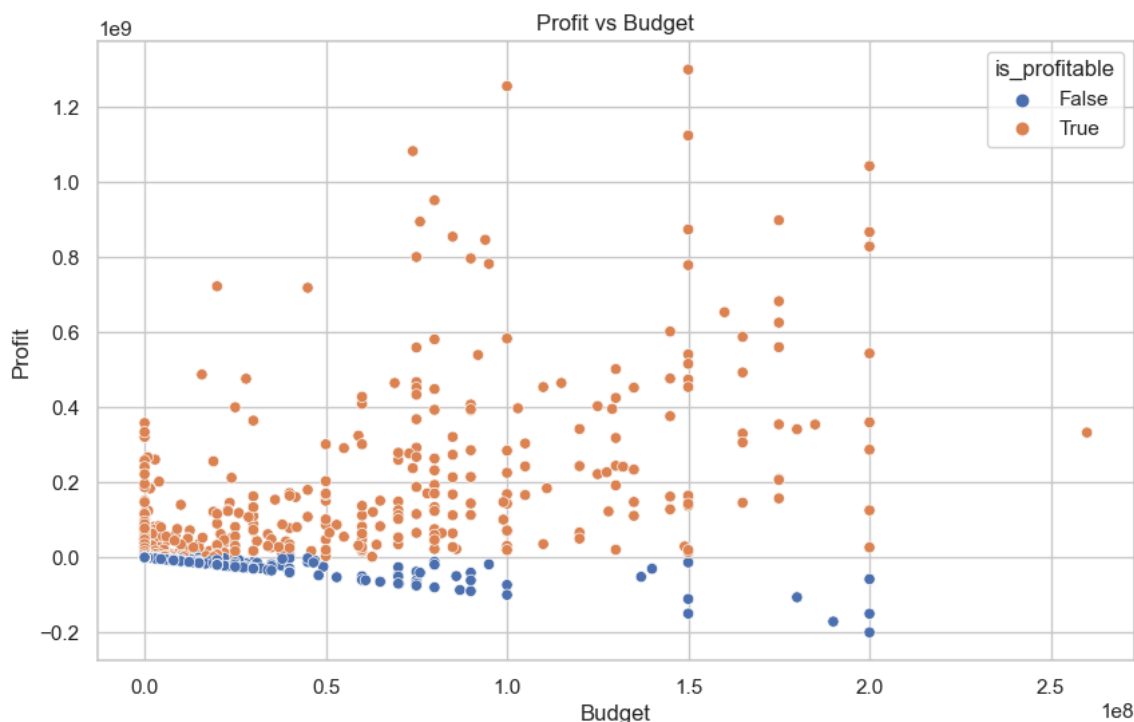
1. **Домінування англійської мови:** найбільша кількість фільмів була випущена англійською мовою. Їх кількість перевищує 25,000, що значно випереджає інші мови. Це вказує на домінування англійської мови у виробництві анімаційних фільмів, оскільки вона є найбільш поширеною у світовій кіноіндустрії та має широку аудиторію.
2. **Популярність японської мови:** японська мова займає друге місце з набагато меншою кількістю фільмів (приблизно 5,000), що свідчить про високий рівень виробництва анімаційних фільмів у Японії, зокрема через популярність аніме.
3. **Французька, російська та іспанська мови:** ці мови займають третє, четверте та п'яте місця відповідно. Кількість цими мовами коливається в межах від 1,000 до 3,000. Це вказує на значний, хоча і менший, внесок країн, що розмовляють цими мовами, у виробництво анімаційних фільмів.
4. **Різноманітність мов:** Хоча англійська домінує, графік також показує, що анімаційні фільми випускаються різними мовами, що дозволяє студіям охопити міжнародну аудиторію та забезпечити культурну різноманітність у контенті.

Таким чином, цей графік демонструє значну перевагу англійської мови у виробництві анімаційних фільмів, а також підкреслює помітний внесок японської, французької, російської та іспанської мов у цій галузі.

### 5.7. Залежність прибутку від бюджету

```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='budget', y='profit', data=data, hue= "is_profitable" )
plt.title('Profit vs Budget')
plt.xlabel('Budget')
plt.ylabel('Profit')
plt.show()
```

Результат:



Візуалізація відображає залежність прибутку від бюджету для анімаційних фільмів, з розподілом на прибуткові (позначені помаранчевими точками) та неприбуткові (сині точки) фільми. Основні спостереження:

1. **Прибутковість при низьких бюджетах:** на графіку видно, що серед фільмів із невеликим бюджетом (до 50 млн\$) є невелика частина прибуткових. Це свідчить про те, що навіть із обмеженими ресурсами можна створювати успішні анімаційні проєкти.
2. **Збільшення бюджету не гарантує прибуток:** хоча деякі фільми з більшим бюджетом (від 100 млн і більше) показують високий прибуток, також є невелика кількість таких, що не є прибутковими. Це підкреслює, що великий бюджет не обов'язково означає комерційний успіх.

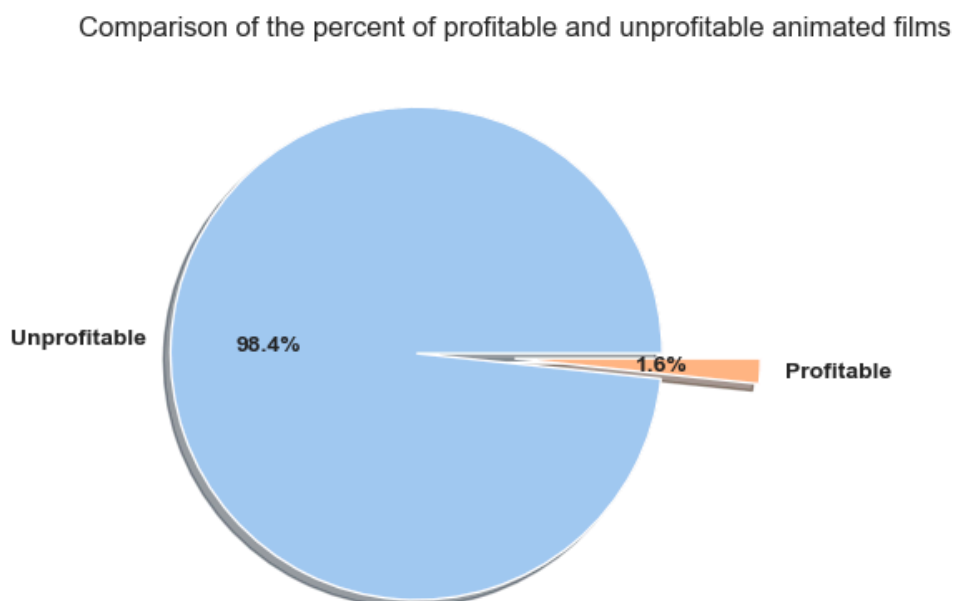
3. **Різкі прибутки при високих бюджетах:** на графіку є декілька точок із дуже високим прибутком (близько 1 млрд), що свідчить про випадки, коли великі інвестиції у фільми призводять до значного прибутку, хоча такі випадки є рідкісними.
4. **Змішана прибутковість серед бюджетів:** У середньому діапазоні бюджетів (50–150 млн) спостерігається як прибутковість, так і збитковість. Це свідчить про те, що фактори, окрім бюджету (як-от маркетинг, сюжет, режисура та цільова аудиторія), відіграють важливу роль у фінансовому успіху фільму.

Отже, дана візуалізація демонструє, що бюджет фільму є важливим фактором для можливого прибутку, проте він не є вирішальним. Низький бюджет може бути достатнім для отримання прибутку, тоді як великий бюджет не гарантує успіху.

### 5.8. Порівняння кількості прибуткових та неприбуткових анімаційних фільмів

```
plt.title('Comparison of the number of profitable and unprofitable
animated films')
plt.pie(data['is_profitable'].value_counts(), labels=['Loss',
'Profit'], autopct="%1.1f%%", shadow=True, textprops={"fontsize": 10,
"weight": "bold"}, colors=sns.color_palette("pastel"), explode = [0.4,
0.0])
plt.show()
plt.show()
```

Результат:

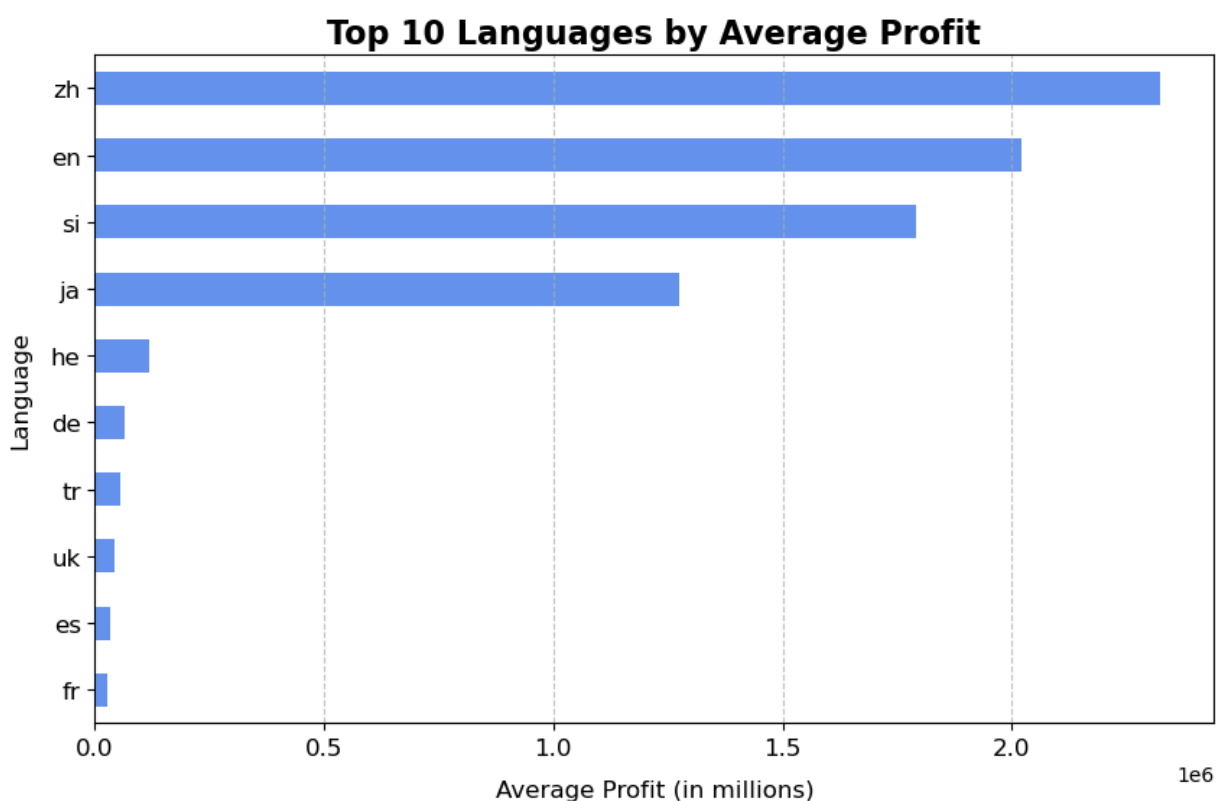


Графік ілюструє розподіл прибуткових та неприбуткових анімаційних фільмів. З нього видно, що значна частка анімаційних фільмів ( $98.4\% \approx 51\,114$ ) є неприбутковою, у той час як лише  $1.6\%$  ( $\approx 831$ ) анімаційних фільмів змогли досягти прибутку. Можна зробити висновки, що створення прибуткового анімаційного фільму є великим викликом. Це підкреслює важливість уважного управління ресурсами та стратегії, які можуть допомогти знизити ризики й підвищити ймовірність успіху проєкту.

### 5.9. 10 мов з найбільшим середнім прибутком

```
average_profit_by_language =
data.groupby('original_language')['profit'].mean().sort_values(ascending=False).head(10)
plt.figure(figsize=(10, 6))
ax = average_profit_by_language.plot(kind='barh', color='cornflowerblue')
ax.set_title("Top 10 Languages by Average Profit", fontsize=16,
weight='bold')
ax.set_xlabel("Average Profit (in millions)", fontsize=12, labelpad=10)
ax.set_ylabel("Language", fontsize=12, labelpad=10)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.grid(axis='x', linestyle='--', alpha=0.7)
plt.show()
```

Результат:



Графік показує 10 мов, фільми на яких мають найбільший середній прибуток. На ньому видно, що:

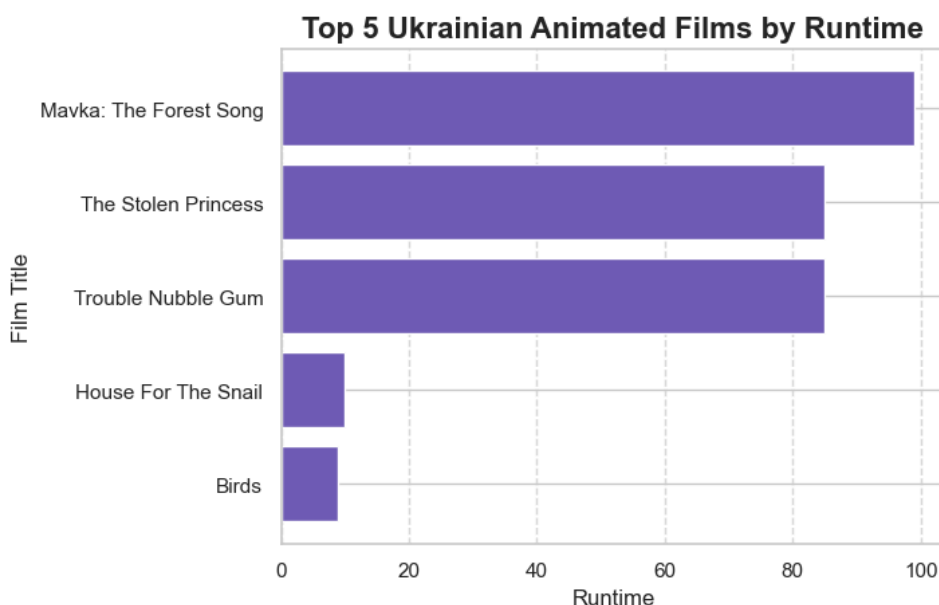
1. **Лідери за середнім прибутком:** китайська (zh) і англійська (en) мови мають найвищий середній прибуток, обидві перевищують 2 мільйони доларів. Це свідчить про великі ринки для фільмів, зокрема в Китаї та англomовних країнах, або про успіх великих кінопроектів, знятих на цих мовах.
2. **Інші мови з високим прибутком:** серед мов з меншим, але все ж значним середнім прибутком - сингальська (si) та японська (ja). Ці мови мають відносно високий середній прибуток, що вказує на наявність успішних фільмів, створених для специфічних ринків, таких як Японія.
3. **Менш прибуткові мови:** інші мови, такі як іврит (he), німецька (de), турецька (tr), українська (uk), іспанська (es) та французька (fr), показують набагато нижчий середній прибуток. Це означає менші розміри ринку або про обмежений успіх фільмів, знятих на цих мовах.



### 5.10. 5 найдовших за тривалістю українських анімаційних фільмів

```
top_ukrainian_animated =
top_ukrainian_animated.sort_values(by='runtime',
ascending=False).head(5)
plt.barh(top_ukrainian_animated['title'],
top_ukrainian_animated['runtime'], color='#705eb5')
plt.xlabel("Runtime", fontsize=12)
plt.ylabel("Film Title", fontsize=12)
plt.title("Top 5 Ukrainian Animated Films by Runtime",
fontsize=16, weight='bold')
plt.gca().invert_yaxis()
plt.grid(axis='x', linestyle='--', alpha=0.7)
```

Результат:



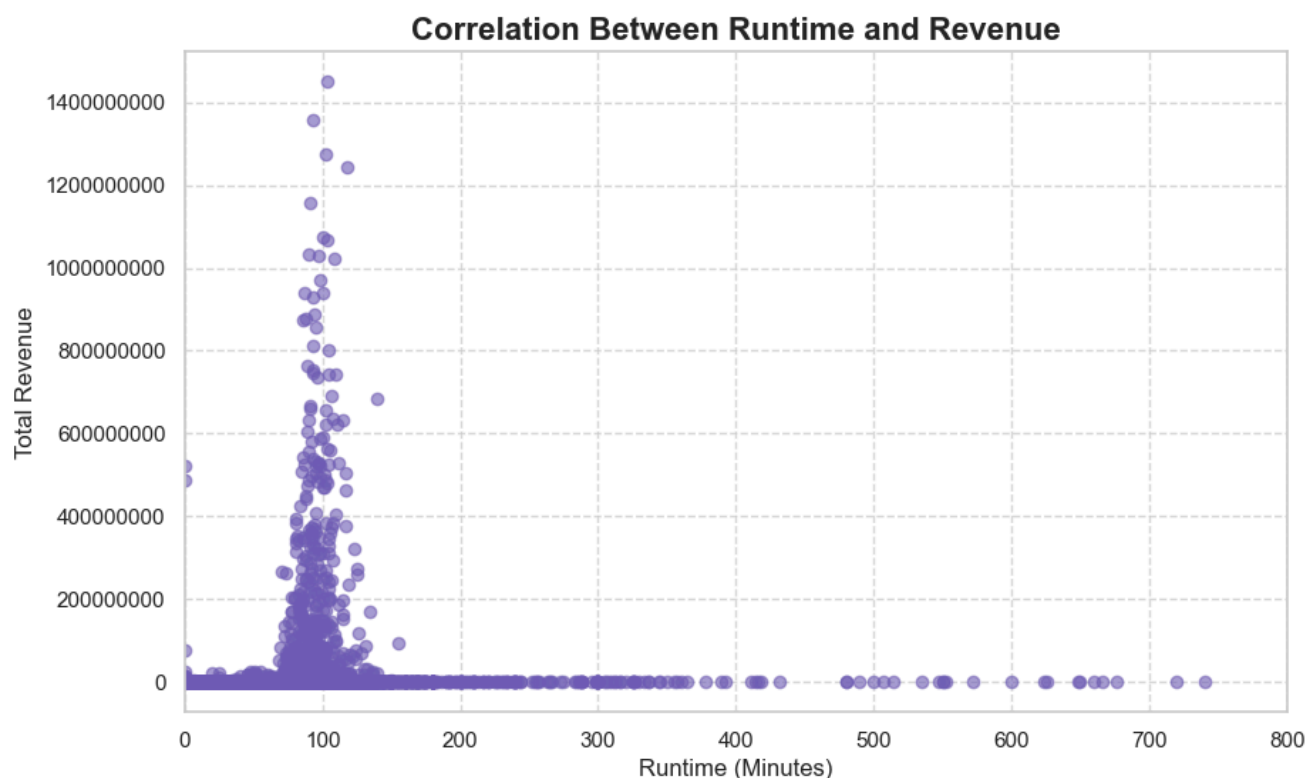
З цієї візуалізації можемо побачити, що весь час українськими студіями було випущено тільки два анімаційних фільми (“[Mavka: The Forest Song](#)” та “[The Stolen Princess](#)”) тривалістю більше, ніж 1 година. Також в датасеті присутня інформація про [Trouble Nubble Gum](#), який є продовженням мультфільму «Як Петрик П'яточкин слоників рахував», але даний анімаційний фільм перебуває у розробці. Інші анімаційні фільми українського виробництва, які представлені в даному датасеті, мають хронометраж до 20 хвилин.

### 5.11. Залежність прибутку від тривалості анімаційного фільму

```
plt.figure(figsize=(10, 6))
plt.scatter(data['runtime'], data['revenue'], color='#705eb5',
            alpha=0.6)
plt.ticklabel_format(style='plain')
plt.xlabel('Runtime (Minutes)', fontsize=12)
plt.ylabel('Total Revenue', fontsize=12)
plt.title('Correlation Between Runtime and Revenue', fontsize=16,
          weight='bold')
plt.xlim(0, 800)

plt.grid(True, linestyle='--', alpha=0.7)
plt.show()
```

Результат:



На графіку зображено зв'язок між тривалістю анімаційних фільмів (у хвилинах) та їхнім загальним прибутком (у доларах). Основні спостереження:

1. Основна кількість фільмів має тривалість до 150 хвилин. Більшість прибуткових фільмів знаходяться в межах цієї тривалості.

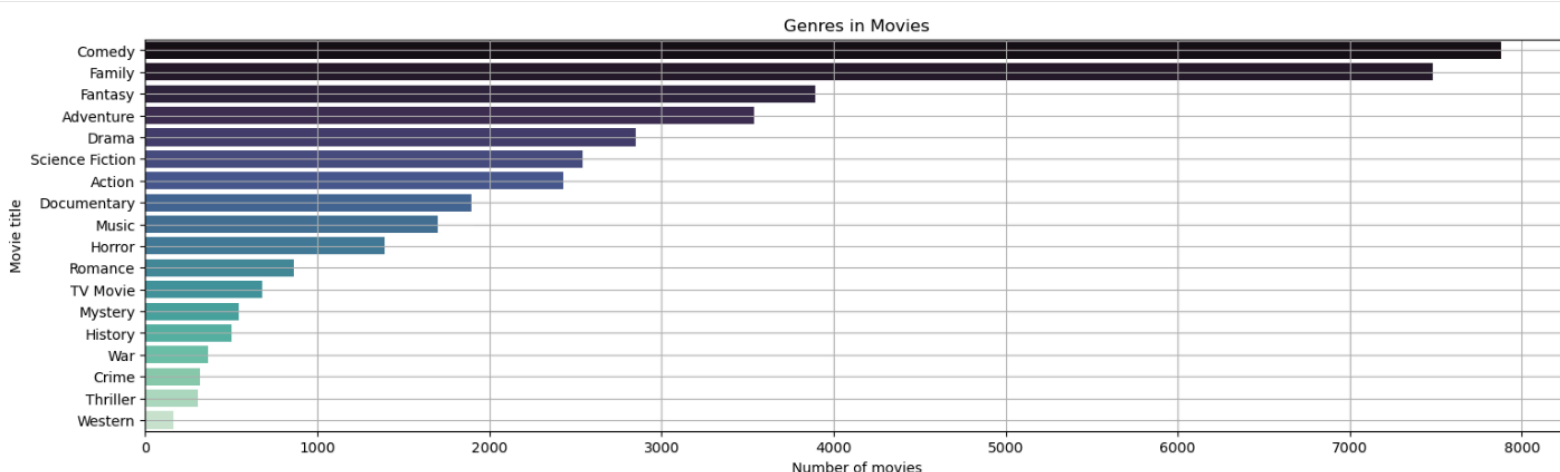
2. Найвищий прибуток зосереджений у фільмів із тривалістю приблизно до 120 хвилин. Фільми з тривалістю понад 150 хвилин, як правило, мають значно менший прибуток.
3. Немає явної залежності між збільшенням тривалості та підвищенням прибутку. Навпаки, фільми з тривалістю понад 200 хвилин практично не демонструють значного прибутку.

Отже, найбільш прибуткові анімаційні фільми, як правило, мають тривалість до 2 годин.

### 5.12. Який анімаційний жанр створюється найбільше

```
plt.figure(figsize=(18,5))
grouped_by_genres_ = grouped_by_genres.iloc[1:]
sns.barplot(y=grouped_by_genres_.index,
x=grouped_by_genres_.values, palette="mako", orient='h')
plt.title("Movie releases By Month")
plt.grid(True)
plt.title("Genres in Movies")
plt.xlabel('Number of movies')
plt.ylabel('Movie title')
plt.show()
```

Результат:



Аналіз графіка жанрової спрямованості анімаційних фільмів показує, що деякі жанри користуються більшою популярністю у виробництві, ніж інші.

Найбільше створюють комедійних анімаційних фільмів (7879), які користуються популярністю завдяки своїй універсальній привабливості для різних вікових груп. Жанр "Сімейний" (7480) займає другу позицію, що підтверджує, що анімація здебільшого орієнтована на сімейну аудиторію. Також популярними є жанри "Фентезі" (3896) та "Пригоди" (3539), які добре поєднуються з анімаційним стилем і дозволяють створювати фантастичні світи, що приваблюють глядачів.

Драма (2849) та наукова фантастика (2540) також зустрічаються в анімаційному кіно, хоча й рідше. Популярними залишаються "Екшн" (2428) та "Документальні" (1897) анімаційні фільми, хоча це й більш специфічні жанри для даного виду кіно.

На противагу популярним жанрам, "Трилер" (306), "Вестерн" (161), а також жанри "Злочин" (320) та "Історія" (499) зустрічаються значно рідше, що може бути пов'язано зі специфікою анімаційного кіно, яке більше орієнтоване на комедійні та фантастичні сюжети.

Отже, можемо зробити такі висновки:

- **Переважаання сімейно-комедійних жанрів:** анімаційні фільми часто орієнтовані на сімейну та молодшу аудиторію, що пояснює домінування комедії та сімейного жанру.
- **Фентезі та пригоди:** ці жанри добре підходять для анімації завдяки можливості створювати чарівні й незвичні світи, що приваблює глядачів усіх вікових категорій.
- **Специфічність жанрів:** жанри, які більшою мірою підходять для реалістичного стилю та драматичної напруги, менш поширені в анімаційному форматі, що підтверджує спрямованість анімаційного кіно на фантазійні та легші теми.

### 5.13. Порівняльний аналіз кількості випущених анімаційних фільмів в період 2001-2010 та 2011-2020 років

```

animation_2000_2010 = data[(data['release_year'] >= 2001) & (data['release_year']
<= 2010)]
animation_2010_2020 = data[(data['release_year'] > 2010) & (data['release_year']
<= 2020)]

count_2000_2010 = animation_2000_2010['release_year'].value_counts().sort_index()
count_2010_2020 = animation_2010_2020['release_year'].value_counts().sort_index()

comparison_years = range(2001, 2011)
y_axis = range(0,2501,250)
comparison_2000s = [count_2000_2010.get(year, 0) for year in comparison_years]
comparison_2010s = [count_2010_2020.get(year + 10, 0) for year in
comparison_years]
comparison_labels = [f"{year} {year + 10}" for year in comparison_years]

fig = plt.figure(figsize=(14, 10))
ax_1 = fig.add_subplot(2, 2, 1)
ax_2 = fig.add_subplot(2, 2, 2)
ax_3 = fig.add_subplot(2, 1, 2)

ax_1.bar(count_2000_2010.index, count_2000_2010.values, color='#5f51b0',width=0.8)
ax_1.set_title('Animated films (2001-2010)')
ax_1.set_xlabel('Year')
ax_1.set_ylabel('Quantity')
ax_1.set_xticks(count_2000_2010.index)
ax_1.set_yticks(y_axis)
ax_2.bar(count_2010_2020.index, count_2010_2020.values, color='#698dcf',width=0.8)
ax_2.set_title('Animated films (2011-2020)')
ax_2.set_xlabel('Year')
ax_2.set_ylabel('Quantity')
ax_2.set_xticks(count_2010_2020.index)
ax_2.set_yticks(y_axis)
fr_len = np.arange(len(comparison_years))
ax_3.bar(fr_len - 0.2, comparison_2000s, width=0.4, color='#5f51b0',
label='2001-2010')
ax_3.bar(fr_len + 0.2, comparison_2010s, width=0.4, color='#698dcf',
label='2010-2020')
ax_3.set_xticks(ticks=fr_len, labels=comparison_labels)
ax_3.set_title('Comparison (2001-2010 vs 2011-2020)')

```

```

ax_3.set_xlabel('Years')
ax_3.set_ylabel('Quantity')
ax_3.legend()

plt.subplots_adjust(wspace=0.2, hspace=0.3)

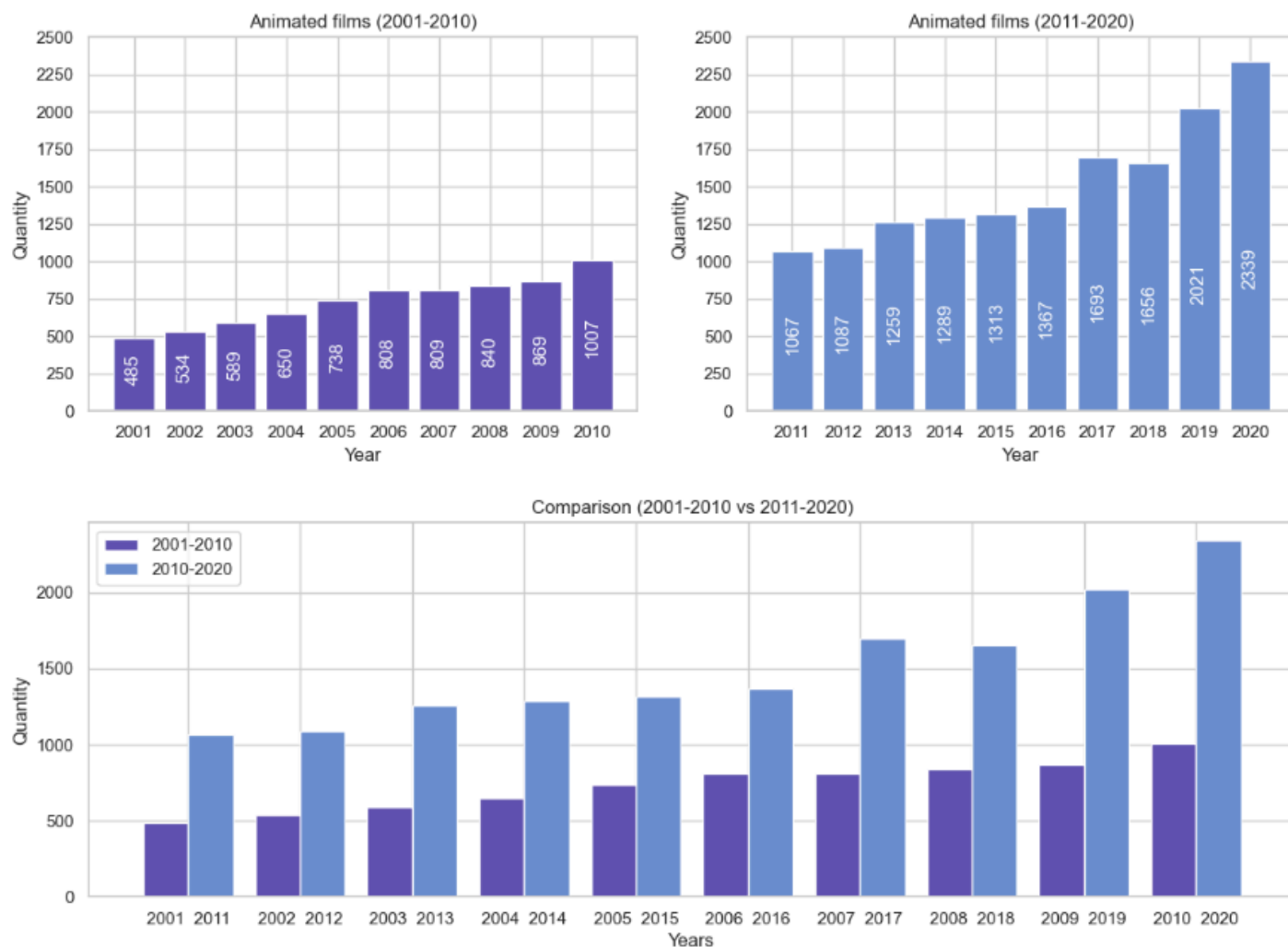
for rect in ax_1.patches:
    height = rect.get_height()
    ax_1.text(rect.get_x() + rect.get_width()/2., 0.35*height, '%d' % int(height),
              ha='center', va='bottom', color='white', rotation=90)

for rect in ax_2.patches:
    height = rect.get_height()
    ax_2.text(rect.get_x() + rect.get_width()/2., 0.35*height, '%d' % int(height),
              ha='center', va='bottom', color='white', rotation=90)

plt.show()

```

Результат:



Графік показує кількість анімаційних фільмів, випущених у періоди 2001-2010 та 2011-2020 років, а також порівняння цих періодів.

Основні спостереження:

1. Ріст кількості фільмів у кожному періоді:
  - a. У період 2001-2010 років спостерігається поступове зростання кількості анімаційних фільмів від 485 у 2001 році до 1007 у 2010 році.
  - b. У період 2011-2020 років кількість фільмів також зростала, починаючи з 1067 у 2011 році і досягнувши 2339 у 2020 році.
2. Суттєве збільшення випуску у 2011-2020 роках:
  - a. Кількість анімаційних фільмів за період 2011-2020 років значно перевищує показники 2001-2010 років. Наприклад, у 2020 році було випущено понад удвічі більше фільмів (2339), ніж у 2010 році (1007).
3. Порівняння двох періодів (нижній графік):
  - a. На нижньому графіку видно, що для кожного року з 2011 по 2020 кількість фільмів значно більша, ніж для відповідного року з 2001 по 2010. Це вказує на загальну тенденцію до збільшення виробництва анімаційних фільмів у 2011-2020 роках.

Отже, можемо зробити висновок, що у період 2011-2020 років відбувся значний ріст кількості анімаційних фільмів порівняно з попереднім десятиріччям (2001-2010). Це пов'язано з технологічним прогресом, збільшенням попиту на анімаційний контент та розширенням платформ для його розповсюдження (стримінгові сервіси тощо).

#### 5.14. Кореляційна матриця основних метрик

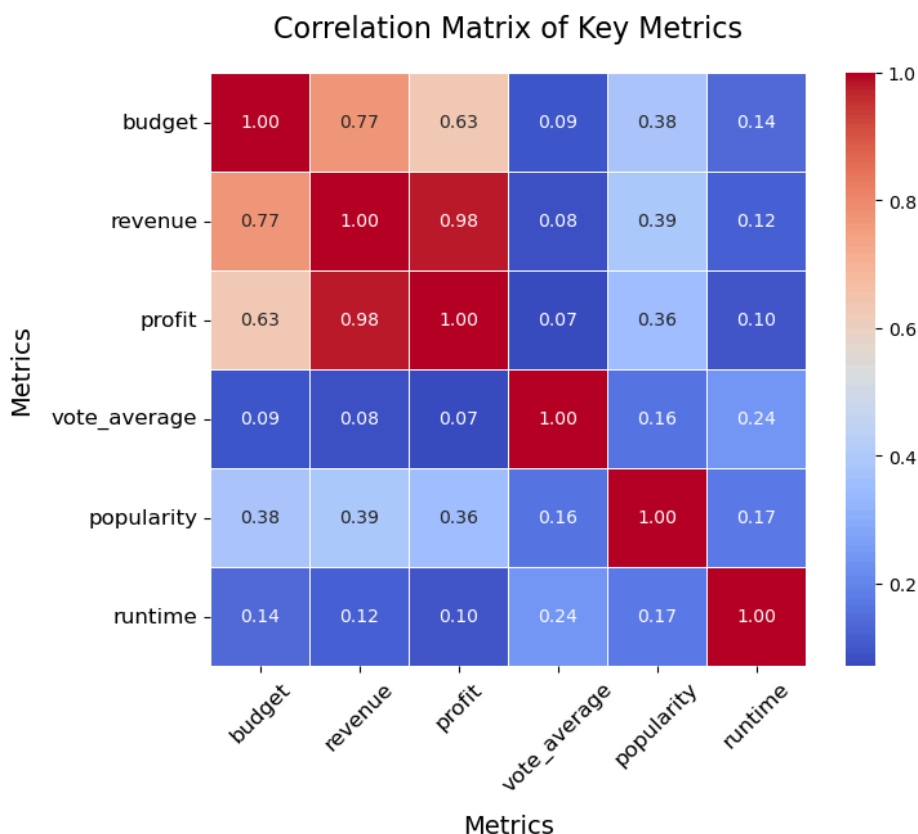
```
corr = data[['budget', 'revenue', 'profit', 'vote_average',
            'popularity', 'runtime']].corr()

plt.figure(figsize=(8, 6))
sns.heatmap(corr, annot=True, fmt=".2f", cmap="coolwarm", cbar=True,
            square=True, linewidths=0.5)

plt.title("Correlation Matrix of Key Metrics", fontsize=16, pad=20)
```

```
plt.xticks(fontsize=12, rotation=45)
plt.yticks(fontsize=12, rotation=0)
plt.xlabel("Metrics", fontsize=14, labelpad=10)
plt.ylabel("Metrics", fontsize=14, labelpad=10)
plt.show()
```

Результат:



Дана візуалізація демонструє матрицю кореляцій між ключовими метриками: бюджет, дохід, прибуток, середній рейтинг та популярність. Кожна комірка відображає силу та напрямок кореляції між двома показниками. Основні спостереження:

### 1. Сильна позитивна кореляція між доходом, прибутком та бюджетом:

Кореляція між бюджетом і доходом становить 0.77, що вказує на сильну позитивну залежність. Це означає, що більший бюджет зазвичай асоціюється з більшим доходом.

Кореляція між доходом і прибутком є майже ідеальною (0.98), що логічно, оскільки прибуток безпосередньо залежить від доходу.



Кореляція між бюджетом і прибутком (0.63) також є значною, що свідчить про те, що більший бюджет зазвичай сприяє вищому прибутку.

## **2. Слабка кореляція середнього рейтингу з іншими показниками:**

Середній рейтинг має низький коефіцієнт кореляції з бюджетом (0.09), доходом (0.08), прибутком (0.07) та популярністю (0.16). Це вказує на те, що рейтинг фільму майже не залежить від цих показників.

## **3. Помірна позитивна кореляція між популярністю та іншими метриками:**

Популярність має помірну кореляцію з бюджетом (0.38), доходом (0.39) і прибутком (0.36). Це означає, що вищий бюджет і прибуток частково сприяють підвищенню популярності, хоча цей зв'язок не є дуже сильним.

Можемо зробити висновок, що бюджет, дохід і прибуток мають тісний взаємозв'язок, що вказує на те, що інвестування більшого бюджету може збільшити дохід та прибуток. Однак середній рейтинг фільму слабо корелює з іншими метриками, що означає, що висока якість (оцінена глядачами) не завжди приводить до більшого доходу чи популярності.

## **Письмове обґрунтування результату**

Результати [дослідження та відповідні графічні візуалізації](#) допомогли визначити основні тенденції та закономірності в анімаційній індустрії з різних аспектів.

- **Країни-виробники:** аналіз країн-виробників анімаційних фільмів показав, що лідером є США, на частку яких припадає значний відсоток продукції. Такі країни, як Японія, Франція, Велика Британія та інші, також відіграють важливу роль, проте у меншій мірі. Це демонструє домінування США на ринку анімаційного кіно, що підкреслено візуалізацією у вигляді кругової діаграми.
- **Виробничі компанії:** серед виробничих компаній виділяються студії ONF/NFB, Soyuzmultfilm, Walt Disney Productions та інші. Це свідчить про значний вплив даних студій на загальний обсяг

продукції, а також їхню роль у формуванні анімаційного ринку. Графік наочно демонструє їхню відносну частку в індустрії.

- **Популярність анімаційних фільмів з роками:** лінійний графік, що відображає середню популярність анімаційних фільмів з роками, показує поступове зростання інтересу до анімації, особливо після 1950-х років, з різким підйомом у 2000-х. Це вказує на еволюцію жанру та його зростаючу актуальність.
- **Прибутковість анімаційних фільмів:** дослідження залежності прибутковості від бюджету показало, що більші бюджети можуть збільшувати ймовірність прибутковості, але це не є гарантією успіху. Точкова діаграма наочно показує, що фільми з високим бюджетом можуть бути як успішними, так і збитковими.
- **Відсоток прибуткових фільмів:** кругова діаграма демонструє, що лише невеликий відсоток (1.6%) анімаційних фільмів є прибутковими. Це свідчить про ризикованість індустрії, де більшість проектів не окуповуються.
- **Прибутковість за мовами:** стовпчаста діаграма, яка відображає 10 мов з найбільшим середнім прибутком, показує, що фільми на китайській (zh) та англійській (en) мовах є найбільш прибутковими. Це підкреслює важливість аудиторії, що говорить цими мовами, для комерційного успіху.
- **Жанри:** дослідження показує, що найбільш популярними жанрами для анімаційних фільмів є комедія, сімейний та фентезі, що відповідає типовим запитам на анімаційне кіно для широкої аудиторії. Відповідний графік наочно показує частку кожного жанру.

Загалом, проведений аналіз та візуалізації допомагають побачити не лише загальні тенденції та закономірності в анімаційній індустрії, але й конкретні фактори, що впливають на прибутковість, популярність і жанрову структуру анімаційного кіно.