

Nir

May 16, 2024

1)

«Google Play Store Apps». 2021. 10000 Google Play Store.

```
[1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[2]: file_path = "/home/margarita/Google-Playstore.csv"
data = pd.read_csv(file_path)
data
```

```
[2]:
```

	App Name \
0	Gakondo
1	Ampere Battery Info
2	Vibook
3	Smart City Trichy Public Service Vehicles 17UC...
4	GROW.me
...	...
2312939	-
2312940	ORU Online
2312941	Data Structure
2312942	Devi Suktam
2312943	Biliyor Musun - Sonsuz Yarış

	App Id	Category	Rating	\
0	com.ishakwe.gakondo	Adventure	0.0	
1	com.webserveis.batteryinfo	Tools	4.4	
2	com.doantiepvien.crm	Productivity	0.0	
3	cst.stJoseph.ug17ucs548	Communication	5.0	
4	com.horodyski.grower	Tools	0.0	
...	
2312939	com.rxsj.ssjj	Role Playing	4.3	
2312940	com.threedream.oruonline	Education	0.0	

2312941	datastructure.appoworld.datastructure	Education	0.0
2312942	ishan.devi.suktam	Music & Audio	3.5
2312943	com.yyazilim.biliyormusun	Trivia	5.0

	Rating	Count	Installs	Minimum Installs	Maximum Installs	Free	\
0		0.0	10+		10.0	15	True
1		64.0	5,000+		5000.0	7662	True
2		0.0	50+		50.0	58	True
3		5.0	10+		10.0	19	True
4		0.0	100+		100.0	478	True
...
2312939		16775.0	100,000+		100000.0	337109	True
2312940		0.0	100+		100.0	430	True
2312941		0.0	100+		100.0	202	True
2312942		8.0	1,000+		1000.0	2635	True
2312943		12.0	100+		100.0	354	True

	Price	...	Developer Website	\
0	0.0	...	https://beniyizibyose.tk/#/	
1	0.0	...	https://webserveis.netlify.app/	
2	0.0	...	NaN	
3	0.0	...	http://www.climatesmarttech.com/	
4	0.0	...	http://www.horodyski.com.pl	
...	
2312939	0.0	...	http://www.4399sy.com.hk/	
2312940	0.0	...	http://www.oru.edu/	
2312941	0.0	...	NaN	
2312942	0.0	...	https://a70f78905.app-ads-txt.com	
2312943	0.0	...	NaN	

	Developer Email	Released	Last Updated	\
0	jean21101999@gmail.com	Feb 26, 2020	Feb 26, 2020	
1	webserveis@gmail.com	May 21, 2020	May 06, 2021	
2	vnacrewit@gmail.com	Aug 9, 2019	Aug 19, 2019	
3	climatesmarttech2@gmail.com	Sep 10, 2018	Oct 13, 2018	
4	rmilekhorodyski@gmail.com	Feb 21, 2020	Nov 12, 2018	
...	
2312939	ssjjcomhk@gmail.com	NaN	Jun 01, 2021	
2312940	3DreamDeveloper@gmail.com	Jan 17, 2018	Feb 02, 2018	
2312941	appoworld.official@gmail.com	Aug 19, 2018	Aug 19, 2018	
2312942	ruchisono@gmail.com	Aug 1, 2016	May 05, 2021	
2312943	yyazilimdevelop@gmail.com	Aug 9, 2019	Aug 19, 2019	

	Content Rating	Privacy Policy	\
0	Everyone	https://beniyizibyose.tk/projects/	
1	Everyone	https://dev4phones.wordpress.com/licencia-de-uso/	
2	Everyone	https://www.vietnamairlines.com/vn/en/terms-an...	

3	Everyone	NaN
4	Everyone	http://www.horodyski.com.pl
...
2312939	Teen	http://a.4399sy.com.hk/user/aggreement
2312940	Everyone	http://www.oru.edu/about-oru/privacy-policy.php
2312941	Everyone	https://appoworld.000webhostapp.com/datastruct...
2312942	Everyone	https://docs.google.com/document/d/1x-9reZuLRX...
2312943	Everyone	https://biliyor-musun-sons.flycricket.io/priva...

	Ad Supported	In App Purchases	Editors Choice	Scraped Time
0	False	False	False	2021-06-15 20:19:35
1	True	False	False	2021-06-15 20:19:35
2	False	False	False	2021-06-15 20:19:35
3	True	False	False	2021-06-15 20:19:35
4	False	False	False	2021-06-15 20:19:35
...
2312939	False	False	False	2021-06-16 12:59:18
2312940	False	False	False	2021-06-16 12:59:19
2312941	False	False	False	2021-06-16 12:59:19
2312942	True	False	False	2021-06-16 12:59:19
2312943	True	False	False	2021-06-16 12:59:19

[2312944 rows x 24 columns]

2)

```
[3]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2312944 entries, 0 to 2312943
Data columns (total 24 columns):
#   Column              Dtype
---  -
0   App Name            object
1   App Id              object
2   Category            object
3   Rating              float64
4   Rating Count        float64
5   Installs            object
6   Minimum Installs    float64
7   Maximum Installs    int64
8   Free                bool
9   Price              float64
10  Currency            object
11  Size                object
12  Minimum Android     object
13  Developer Id        object
14  Developer Website   object
```

```

15 Developer Email    object
16 Released          object
17 Last Updated      object
18 Content Rating    object
19 Privacy Policy    object
20 Ad Supported       bool
21 In App Purchases  bool
22 Editors Choice     bool
23 Scraped Time       object
dtypes: bool(4), float64(4), int64(1), object(15)
memory usage: 361.8+ MB

```

:

```
[4]: data.shape
```

```
[4]: (2312944, 24)
```

:

```
[5]: data.describe()
```

```

[5]:
      Rating  Rating Count  Minimum Installs  Maximum Installs \
count  2.290061e+06  2.290061e+06      2.312837e+06      2.312944e+06
mean    2.203152e+00  2.864839e+03      1.834452e+05      3.202017e+05
std     2.106223e+00  2.121626e+05      1.513144e+07      2.355495e+07
min     0.000000e+00  0.000000e+00      0.000000e+00      0.000000e+00
25%     0.000000e+00  0.000000e+00      5.000000e+01      8.400000e+01
50%     2.900000e+00  6.000000e+00      5.000000e+02      6.950000e+02
75%     4.300000e+00  4.200000e+01      5.000000e+03      7.354000e+03
max     5.000000e+00  1.385576e+08      1.000000e+10      1.205763e+10

      Price
count  2.312944e+06
mean    1.034992e-01
std     2.633127e+00
min     0.000000e+00
25%     0.000000e+00
50%     0.000000e+00
75%     0.000000e+00
max     4.000000e+02

```

10 :

```
[6]: data.head(10)
```

```

[6]:
      App Name \
0      Gakondo
1  Ampere Battery Info

```

2	Vibook
3	Smart City Trichy Public Service Vehicles 17UC...
4	GROW.me
5	IMOCCI
6	unlimited 4G data prank free app
7	The Everyday Calendar
8	WhatsOpen
9	Neon 3d Iron Tech Keyboard Theme

	App Id	Category	Rating \
0	com.ishakwe.gakondo	Adventure	0.0
1	com.webserveis.batteryinfo	Tools	4.4
2	com.doantiepvien.crm	Productivity	0.0
3	cst.stJoseph.ug17ucs548	Communication	5.0
4	com.horodyski.grower	Tools	0.0
5	com.imocci	Social	0.0
6	getfreedata.superfatiza.unlimitedjiodataprank	Libraries & Demo	4.5
7	com.mozaix.simoneboard	Lifestyle	2.0
8	com.whatsopen.app	Communication	0.0
9	com.ikeyboard.theme.neon_3d.iron.tech	Personalization	4.7

	Rating	Count	Installs	Minimum Installs	Maximum Installs	Free	Price \
0	0.0	10+	10.0	15	True	0.0	
1	64.0	5,000+	5000.0	7662	True	0.0	
2	0.0	50+	50.0	58	True	0.0	
3	5.0	10+	10.0	19	True	0.0	
4	0.0	100+	100.0	478	True	0.0	
5	0.0	50+	50.0	89	True	0.0	
6	12.0	1,000+	1000.0	2567	True	0.0	
7	39.0	500+	500.0	702	True	0.0	
8	0.0	10+	10.0	18	True	0.0	
9	820.0	50,000+	50000.0	62433	True	0.0	

	Developer Website	Developer Email \
0	https://beniyizibyose.tk/#/	jean21101999@gmail.com
1	https://webserveis.netlify.app/	webserveis@gmail.com
2	NaN	vnacrewit@gmail.com
3	http://www.climatesmarttech.com/	climatesmarttech2@gmail.com
4	http://www.horodyski.com.pl	rmilekhorodyski@gmail.com
5	http://www.imocci.com	info@imocci.com
6	NaN	aitomgharfatimezzahra@gmail.com
7	NaN	elementuser03@gmail.com
8	http://yilvermolinah.com	yilver.mh1996@gmail.com
9	https://trendyteme888-31139.web.app	trendyteme.888@gmail.com

	Released	Last Updated	Content Rating \
0	Feb 26, 2020	Feb 26, 2020	Everyone

1	May 21, 2020	May 06, 2021	Everyone
2	Aug 9, 2019	Aug 19, 2019	Everyone
3	Sep 10, 2018	Oct 13, 2018	Everyone
4	Feb 21, 2020	Nov 12, 2018	Everyone
5	Dec 24, 2018	Dec 20, 2019	Teen
6	Sep 23, 2019	Sep 27, 2019	Everyone
7	Jun 21, 2019	Jun 21, 2019	Everyone
8	NaN	Dec 07, 2018	Teen
9	Sep 22, 2019	Oct 07, 2020	Everyone

	Privacy Policy	Ad Supported	\
0	https://beniyizibyose.tk/projects/	False	
1	https://dev4phones.wordpress.com/licencia-de-uso/	True	
2	https://www.vietnamairlines.com/vn/en/terms-an...	False	
3	NaN	True	
4	http://www.horodyski.com.pl	False	
5	https://www.imocci.com/wp-content/uploads/2018...	False	
6	https://sites.google.com/view/unlimited4gdatap...	True	
7	https://www.freeprivacypolicy.com/privacy/view...	False	
8	http://elcafedelamanana.yilvermolinah.com/poli...	False	
9	http://bit.ly/EmojiThemePro	True	

	In App Purchases	Editors Choice	Scraped Time
0	False	False	2021-06-15 20:19:35
1	False	False	2021-06-15 20:19:35
2	False	False	2021-06-15 20:19:35
3	False	False	2021-06-15 20:19:35
4	False	False	2021-06-15 20:19:35
5	True	False	2021-06-15 20:19:35
6	False	False	2021-06-15 20:19:35
7	False	False	2021-06-15 20:19:35
8	False	False	2021-06-15 20:19:35
9	False	False	2021-06-15 20:19:35

[10 rows x 24 columns]

3)

```
[7]: data.isna().sum()
```

```
[7]: App Name          5
     App Id            0
     Category          0
     Rating            22883
     Rating Count      22883
     Installs          107
     Minimum Installs  107
```

Maximum Installs	0
Free	0
Price	0
Currency	135
Size	196
Minimum Android	6530
Developer Id	33
Developer Website	760835
Developer Email	31
Released	71053
Last Updated	0
Content Rating	0
Privacy Policy	420953
Ad Supported	0
In App Purchases	0
Editors Choice	0
Scraped Time	0
dtype:	int64

```
[8]: data.duplicated().sum()
```

```
[8]: 0
```

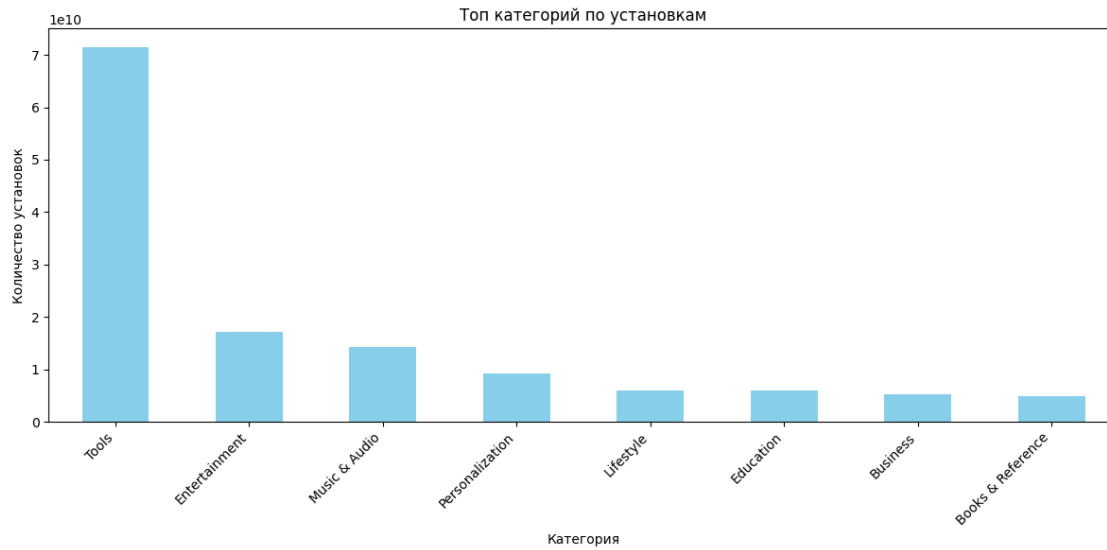
```
[9]: data['Category'].value_counts()
```

```
[9]: Category
Education                241090
Music & Audio             154906
Tools                    143988
Business                  143771
Entertainment             138276
Lifestyle                 118331
Books & Reference         116728
Personalization           89210
Health & Fitness          83510
Productivity              79698
Shopping                  75256
Food & Drink              73927
Travel & Local            67288
Finance                   65466
Arcade                    53792
Puzzle                    51168
Casual                    50813
Communication             48167
Sports                    47483
Social                    44734
News & Magazines          42807
```

Photography	35552
Medical	32065
Action	27555
Maps & Navigation	26722
Simulation	23282
Adventure	23203
Educational	21308
Art & Design	18539
Auto & Vehicles	18280
House & Home	14369
Video Players & Editors	14015
Events	12841
Trivia	11795
Beauty	11772
Board	10588
Racing	10362
Role Playing	10034
Word	8630
Strategy	8526
Card	8179
Weather	7246
Dating	6524
Libraries & Demo	5198
Casino	5076
Music	4202
Parenting	3810
Comics	2862

Name: count, dtype: int64

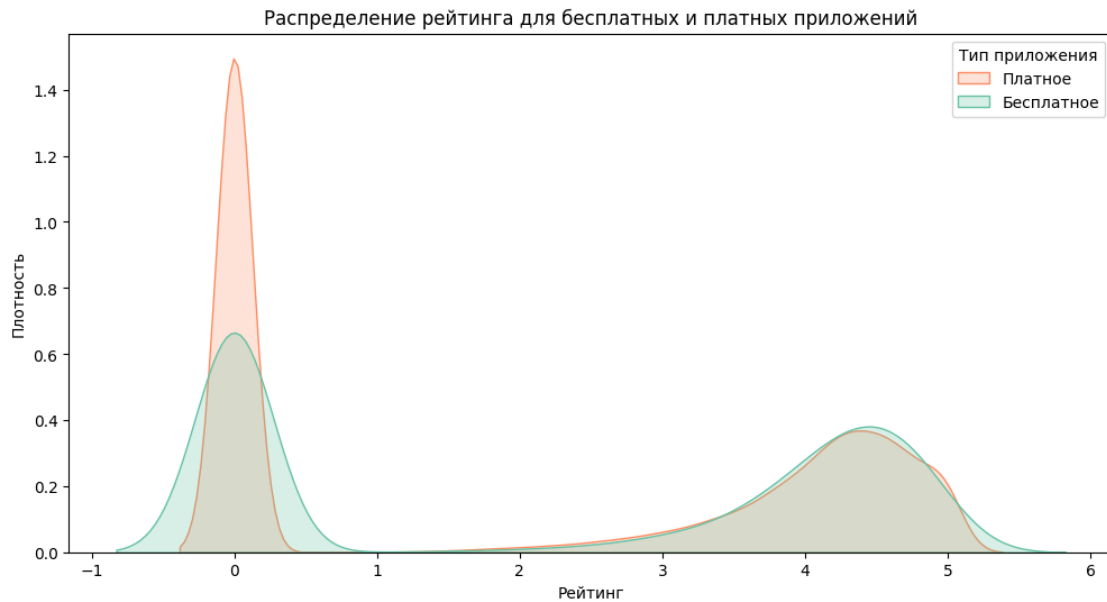
```
[10]: top_8_list = [
    'Education', 'Music & Audio', 'Tools', 'Business',
    'Entertainment', 'Lifestyle', 'Books & Reference',
    'Personalization'
]
top = data[data['Category'].isin(top_8_list)].reset_index(drop=True)
category_installs = top.groupby('Category')['Minimum Installs'].sum().
    ↪sort_values(ascending=False)
plt.figure(figsize=(12, 6))
category_installs.plot(kind='bar', color='skyblue')
plt.title('')
plt.xlabel('')
plt.ylabel('')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

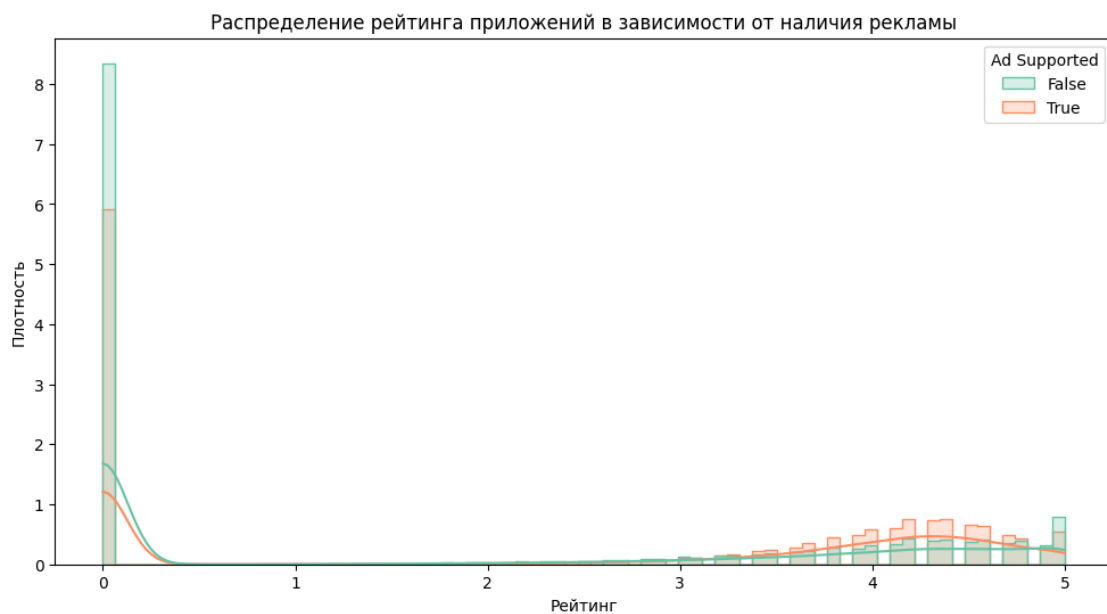
```
[11]: top_categories = data['Category'].value_counts().head(10).index

#
filtered_data = data[data['Category'].isin(top_categories)]

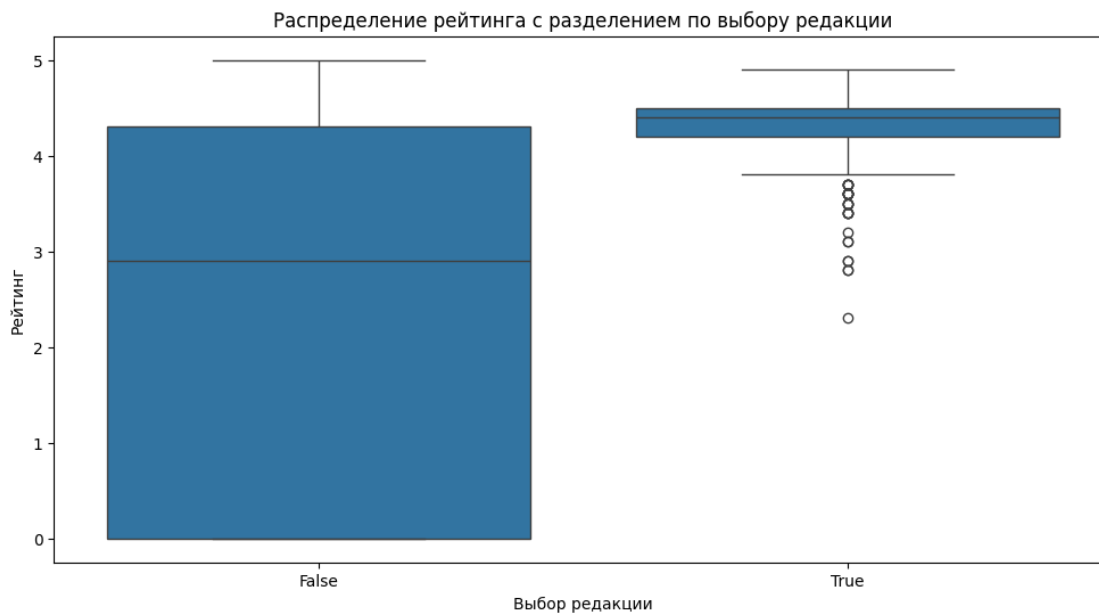
# KDE
plt.figure(figsize=(12, 6))
sns.kdeplot(data=filtered_data, x='Rating', hue='Free', common_norm=False,
            fill=True, palette='Set2')
plt.title('')
plt.xlabel('')
plt.ylabel('')
plt.legend(title='', labels=['', ''])
plt.show()
```



```
[12]: plt.figure(figsize=(12, 6))
sns.histplot(data, x='Rating', hue='Ad Supported', kde=True, palette='Set2',
             element='step', stat='density', common_norm=False)
plt.title('')
plt.xlabel('')
plt.ylabel('')
plt.show()
```

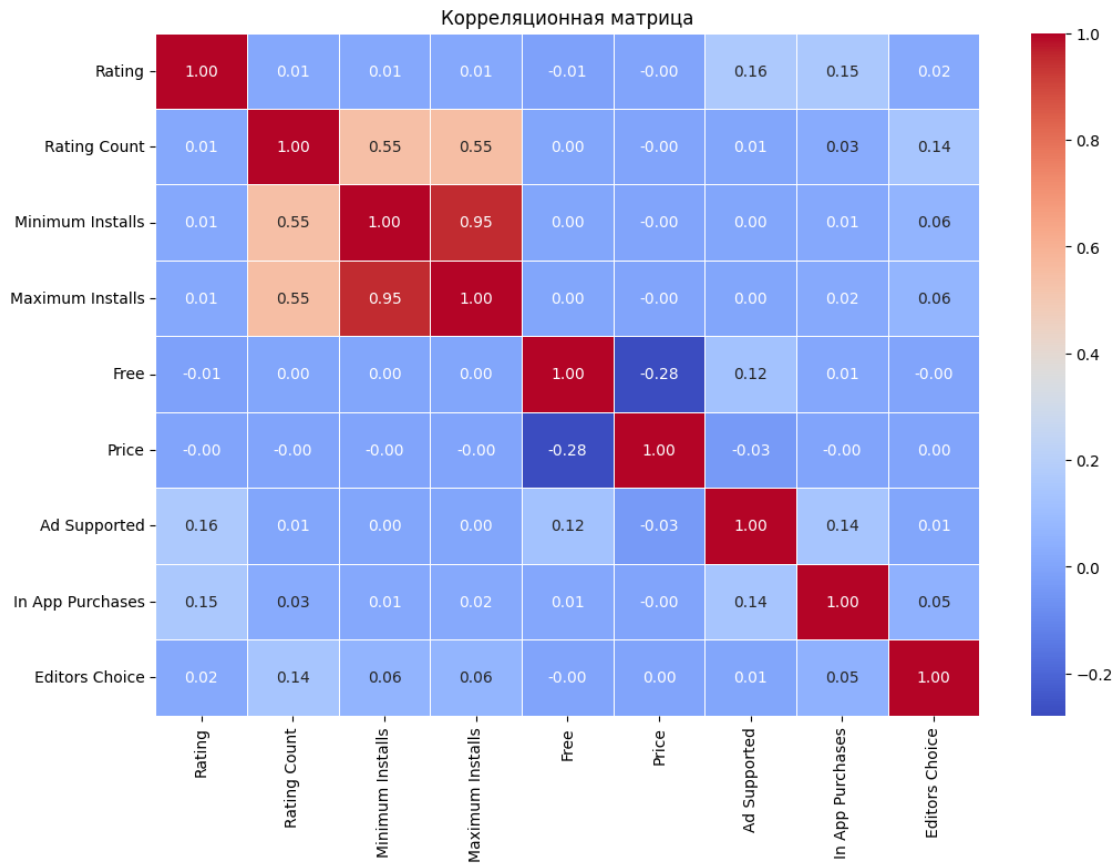


```
[13]: plt.figure(figsize=(12, 6))
sns.boxplot(x='Editors Choice', y='Rating', data=data)
plt.title('')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



4)

```
[14]: data_numeric = data.select_dtypes(include=['float64', 'int64', 'bool'])
correlation_matrix = data_numeric.corr()
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f",
            linewidths=.5)
plt.title("")
plt.show()
```



[]: