

```
In [8]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os

# Настройка внешнего вида графиков
sns.set(style="whitegrid")

# Получаем текущую директорию и формируем путь к файлу
current_dir = os.getcwd()
file_path = os.path.join(current_dir, 'googleplaystore.csv')

# Загрузка DataFrame и вывод первых пяти строк
df = pd.read_csv(file_path)
print(df.head())

# Количество строк и столбцов в DataFrame
print("\nКоличество строк и столбцов:", df.shape)

# Основная информация о DataFrame и анализ пропусков
df.info()

# Удаление или замена пропусков
df['Rating'].fillna(df['Rating'].median(), inplace=True)
df.dropna(subset=['Genres'], inplace=True)

# Проверка на пропуски после обработки
print("\nКоличество пропусков после обработки:")
print(df.isnull().sum())

# Убедимся, что столбец 'Price' в строковом формате и очистим его от симв
df['Price'] = df['Price'].astype(str).str.replace('$', '', regex=False).s
df['Price'] = pd.to_numeric(df['Price'], errors='coerce')
print("\nМинимальная цена:", df['Price'].min())
print("Максимальная цена:", df['Price'].max())

# Убеждаемся, что 'Reviews' в целочисленном формате
df['Reviews'] = pd.to_numeric(df['Reviews'], errors='coerce').fillna(0).a
print("\nМедиана рейтинга:", df['Rating'].median())
print("Средний рейтинг:", df['Rating'].mean())
print("Медиана отзывов:", df['Reviews'].median())
print("Среднее количество отзывов:", df['Reviews'].mean())

# Уникальные значения в Genres
print("\nУникальные значения в Genres:")
print(df['Genres'].unique())

# Сгруппировать по Genres и посчитать средний и медианный рейтинг
grouped_df = df.groupby('Genres')['Rating'].agg(['mean', 'median']).reset
print("\nСредний и медианный рейтинг по жанрам:")
print(grouped_df)

# Жанры с медианным рейтингом больше 4.5
print("\nЖанры с медианным рейтингом больше 4.5:")
print(grouped_df[grouped_df['median'] > 4.5])
```

```

# Удаление дубликатов
initial_duplicates = df.duplicated().sum()
df.drop_duplicates(inplace=True)
print("\nКоличество дубликатов до удаления:", initial_duplicates)
print("Количество дубликатов после удаления:", df.duplicated().sum())

# Количество приложений в каждом жанре
print("\nКоличество приложений в каждом жанре (ТОП-10):")
print(df['Genres'].value_counts().head(10))

# Визуализации
# 1. Горизонтальный столбчатый график для ТОП-10 жанров по количеству при
plt.figure(figsize=(12, 8))
top_genres = df['Genres'].value_counts().head(10)
sns.barplot(x=top_genres.values, y=top_genres.index)
plt.title('ТОП-10 жанров по количеству приложений')
plt.xlabel('Количество приложений')
plt.ylabel('Жанры')
plt.show()

```

	App	Category	Rati
ng \			
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	
4.1			
1	Coloring book moana	ART_AND_DESIGN	
3.9			
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	
4.7			
3	Sketch – Draw & Paint	ART_AND_DESIGN	
4.5			
4	Pixel Draw – Number Art Coloring Book	ART_AND_DESIGN	
4.3			

	Reviews	Size	Installs	Type	Price	Content Rating	\
0	159	19M	10,000+	Free	0.0	Everyone	
1	967	14M	500,000+	Free	0.0	Everyone	
2	87510	8.7M	5,000,000+	Free	0.0	Everyone	
3	215644	25M	50,000,000+	Free	0.0	Teen	
4	967	2.8M	100,000+	Free	0.0	Everyone	

	Genres	Last Updated	Current Ver	\
0	Art & Design	January 7, 2018	1.0.0	
1	Art & Design;Pretend Play	January 15, 2018	2.0.0	
2	Art & Design	August 1, 2018	1.2.4	
3	Art & Design	June 8, 2018	Varies with device	
4	Art & Design;Creativity	June 20, 2018	1.1	

	Android Ver
0	4.0.3 and up
1	4.0.3 and up
2	4.0.3 and up
3	4.2 and up
4	4.4 and up

Количество строк и столбцов: (10841, 13)
 <class 'pandas.core.frame.DataFrame'>

RangeIndex: 10841 entries, 0 to 10840

Data columns (total 13 columns):

#	Column	Non-Null Count	Dtype
0	App	10841 non-null	object
1	Category	10841 non-null	object
2	Rating	9367 non-null	float64
3	Reviews	10841 non-null	object
4	Size	10841 non-null	object
5	Installs	10841 non-null	object
6	Type	10840 non-null	object
7	Price	10841 non-null	float64
8	Content Rating	10840 non-null	object
9	Genres	10841 non-null	object
10	Last Updated	10841 non-null	object
11	Current Ver	10833 non-null	object
12	Android Ver	10838 non-null	object

dtypes: float64(2), object(11)

memory usage: 1.1+ MB

Количество пропусков после обработки:

App	0
Category	0
Rating	0
Reviews	0
Size	0
Installs	0
Type	1
Price	0
Content Rating	1
Genres	0
Last Updated	0
Current Ver	8
Android Ver	3

dtype: int64

Минимальная цена: 0.0

Максимальная цена: 400.0

Медиана рейтинга: 4.3

Средний рейтинг: 4.2078406051102295

Медиана отзывов: 2094.0

Среднее количество отзывов: 444111.92629831197

Уникальные значения в Genres:

['Art & Design' 'Art & Design;Pretend Play' 'Art & Design;Creativity'
'Art & Design;Action & Adventure' 'Auto & Vehicles' 'Beauty'
'Books & Reference' 'Business' 'Comics' 'Comics;Creativity'
'Communication' 'Dating' 'Education;Education' 'Education'
'Education;Creativity' 'Education;Music & Video'
'Education;Action & Adventure' 'Education;Pretend Play'
'Education;Brain Games' 'Entertainment' 'Entertainment;Music & Video'
'Entertainment;Brain Games' 'Entertainment;Creativity' 'Events' 'Finance'
'Food & Drink' 'Health & Fitness' 'House & Home' 'Libraries & Demo'
'Lifestyle' 'Lifestyle;Pretend Play' 'Adventure;Action & Adventure'
'Arcade' 'Casual' 'Card' 'Casual;Pretend Play' 'Action' 'Strategy'
'Puzzle' 'Sports' 'Music' 'Word' 'Racing' 'Casual;Creativity']

```
'Casual;Action & Adventure' 'Simulation' 'Adventure' 'Board' 'Trivia'
'Role Playing' 'Simulation;Education' 'Action;Action & Adventure'
'Casual;Brain Games' 'Simulation;Action & Adventure'
'Educational;Creativity' 'Puzzle;Brain Games' 'Educational;Education'
'Card;Brain Games' 'Educational;Brain Games' 'Educational;Pretend Play'
'Entertainment;Education' 'Casual;Education' 'Music;Music & Video'
'Racing;Action & Adventure' 'Arcade;Pretend Play'
'Role Playing;Action & Adventure' 'Simulation;Pretend Play'
'Puzzle;Creativity' 'Sports;Action & Adventure'
'Educational;Action & Adventure' 'Arcade;Action & Adventure'
'Entertainment;Action & Adventure' 'Puzzle;Action & Adventure'
'Strategy;Action & Adventure' 'Music & Audio;Music & Video'
'Health & Fitness;Education' 'Adventure;Education' 'Board;Brain Games'
'Board;Action & Adventure' 'Board;Pretend Play' 'Casual;Music & Video'
'Role Playing;Pretend Play' 'Entertainment;Pretend Play'
'Video Players & Editors;Creativity' 'Card;Action & Adventure' 'Medical'
'Social' 'Shopping' 'Photography' 'Travel & Local'
'Travel & Local;Action & Adventure' 'Tools' 'Tools;Education'
'Personalization' 'Productivity' 'Parenting' 'Parenting;Music & Video'
'Parenting;Education' 'Parenting;Brain Games' 'Weather'
'Video Players & Editors' 'Video Players & Editors;Music & Video'
'News & Magazines' 'Maps & Navigation'
'Health & Fitness;Action & Adventure' 'Educational' 'Casino'
'Adventure;Brain Games' 'Trivia;Education' 'Lifestyle;Education'
'Books & Reference;Creativity' 'Books & Reference;Education'
'Puzzle;Education' 'Role Playing;Education' 'Role Playing;Brain Games'
'Strategy;Education' 'Racing;Pretend Play' 'Communication;Creativity'
'February 11, 2018' 'Strategy;Creativity']
```

Средний и медианный рейтинг по жанрам:

	Genres	mean	median
0	Action	4.285753	4.3
1	Action;Action & Adventure	4.311765	4.3
2	Adventure	4.184000	4.3
3	Adventure;Action & Adventure	4.423077	4.5
4	Adventure;Brain Games	4.600000	4.6
..
115	Video Players & Editors	4.084393	4.3
116	Video Players & Editors;Creativity	4.100000	4.1
117	Video Players & Editors;Music & Video	4.000000	4.0
118	Weather	4.248780	4.3
119	Word	4.406897	4.4

[120 rows x 3 columns]

Жанры с медианным рейтингом больше 4.5:

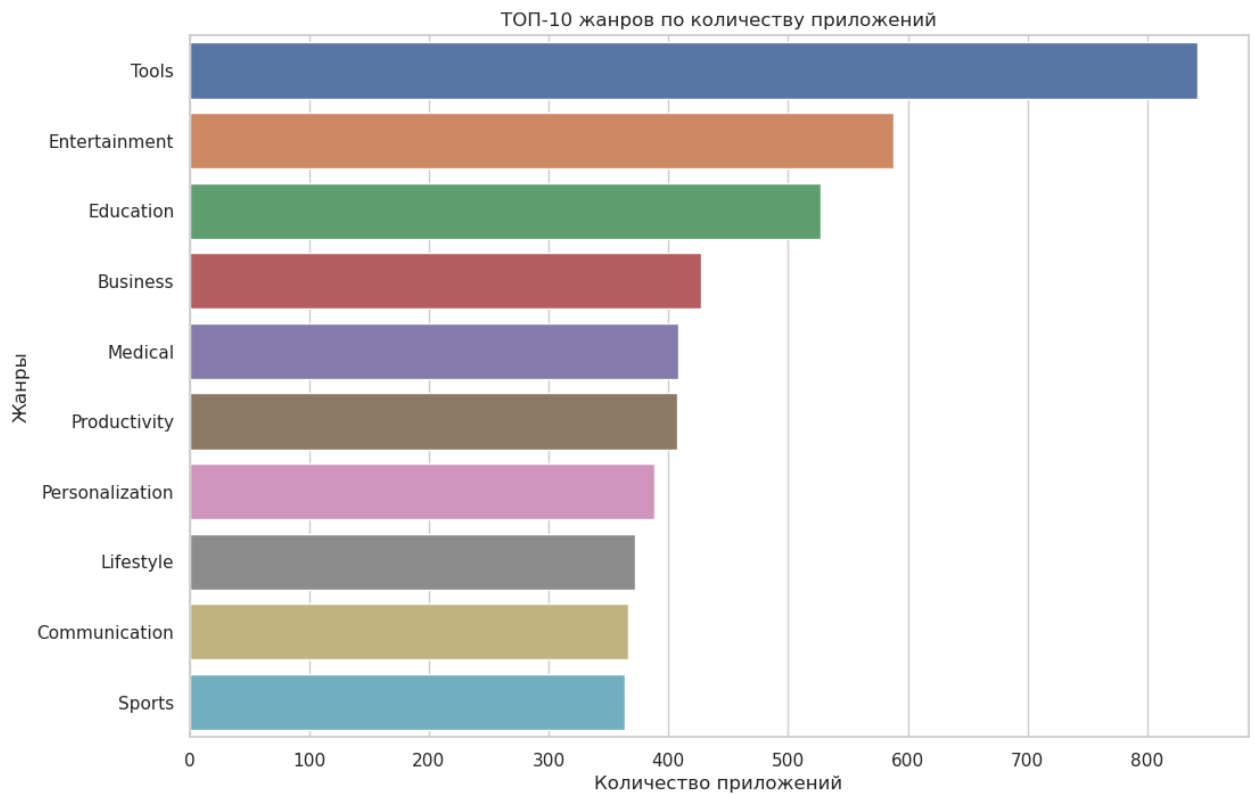
	Genres	mean	median
4	Adventure;Brain Games	4.600000	4.6
11	Art & Design;Creativity	4.400000	4.7
18	Board;Pretend Play	4.800000	4.8
35	Comics;Creativity	4.800000	4.8
55	Entertainment;Creativity	4.533333	4.6
60	February 11, 2018	19.000000	19.0
65	Health & Fitness;Education	4.700000	4.7
75	Music;Music & Video	4.533333	4.6
88	Puzzle;Education	4.600000	4.6
106	Strategy;Action & Adventure	4.600000	4.6

Количество дубликатов до удаления: 483
Количество дубликатов после удаления: 0

Количество приложений в каждом жанре (ТОП-10):

Tools	842
Entertainment	588
Education	527
Business	427
Medical	408
Productivity	407
Personalization	388
Lifestyle	372
Communication	366
Sports	364

Name: Genres, dtype: int64



In []: