

Извлечение данных из интернет-магазина Steam

Егорова Лада Владимировна,
444 группа

st068843@student.spbu.ru

14 апреля 2022 г.

Введение

- Steam — один из крупнейших продавцов компьютерных игр
- Он предоставляет множество фильтров, по которым можно найти всё, что требуется



Инди

- Инди-игра — компьютерная игра, созданная отдельным разработчиком или небольшим коллективом без финансовой поддержки издателя компьютерных игр (Википедия)
- Сначала найдём игры по тегу “Инди” и по исключению этого тега



Данные для изучения

	Название	Цена	Метки	Тип	Дата выхода	Отзывы	Рейтинг	Число отзывов
0	Phasmophobia	309.0	['Horror', 'Online Co-Op', 'Multiplayer', 'Psy...	indie	2020-09-18	Overwhelmingly Positive	97%	339,441
1	Rust	3000.0	['Survival', 'Crafting', 'Multiplayer', 'Open ...	indie	2018-02-08	Very Positive	87%	620,455
2	DayZ	1199.0	['Survival', 'Zombies', 'Multiplayer', 'Open W...	indie	2018-12-13	Mostly Positive	71%	249,404
3	ARK: Survival Evolved	169.0	['Open World Survival Craft', 'Survival', 'Ope...	indie	2017-08-27	Very Positive	82%	410,820
4	Risk of Rain 2	799.0	['Third-Person Shooter', 'Action Roguelike', '...	indie	2020-08-11	Overwhelmingly Positive	96%	130,405
...
11989	Mystery Case Files: The Countess Collector's E...	259.0	['Adventure', 'Casual', 'Hidden Object']	AAA	2018-11-27	Mixed	66%	15
11990	Bunker Defense	465.0	['Early Access', 'Strategy', 'Tower Defense', ...	AAA	2020-10-14	Positive	93%	16
11991	Be Hero	299.0	['Early Access', 'RPG', 'Farming Sim', 'Crafti...	AAA	2021-03-30	Positive	93%	29
11992	Rift Coaster HD Remastered VR	29.0	['Casual', 'Simulation', 'VR']	AAA	2017-08-09	Mixed	40%	76
11993	Reverie - A Heroes Tale	82.0	['RPG']	AAA	2018-03-21	Mostly Positive	75%	12

11947 rows × 8 columns

Выдвигаемые гипотезы

- Средняя цена игры в инди-сегменте ниже средней цены AAA игр
- AAA игры составляют большую часть среди самых дорогих
- Игры с рейтингом 100% имеют, как правило, очень мало отзывов
- Среди всех игр с меткой «Пиксельная графика» большая часть будет принадлежать к инди



Цены

Средняя цена игры

```
1 df[['Цена']].apply(np.median)
```

```
Цена    259.0  
dtype: float64
```

```
1 df[df['Тип'] == 'indie']['Цена'].apply(np.median)
```

```
Цена    259.0  
dtype: float64
```

```
1 df[df['Тип'] == 'AAA']['Цена'].apply(np.median)
```

```
Цена    259.0  
dtype: float64
```



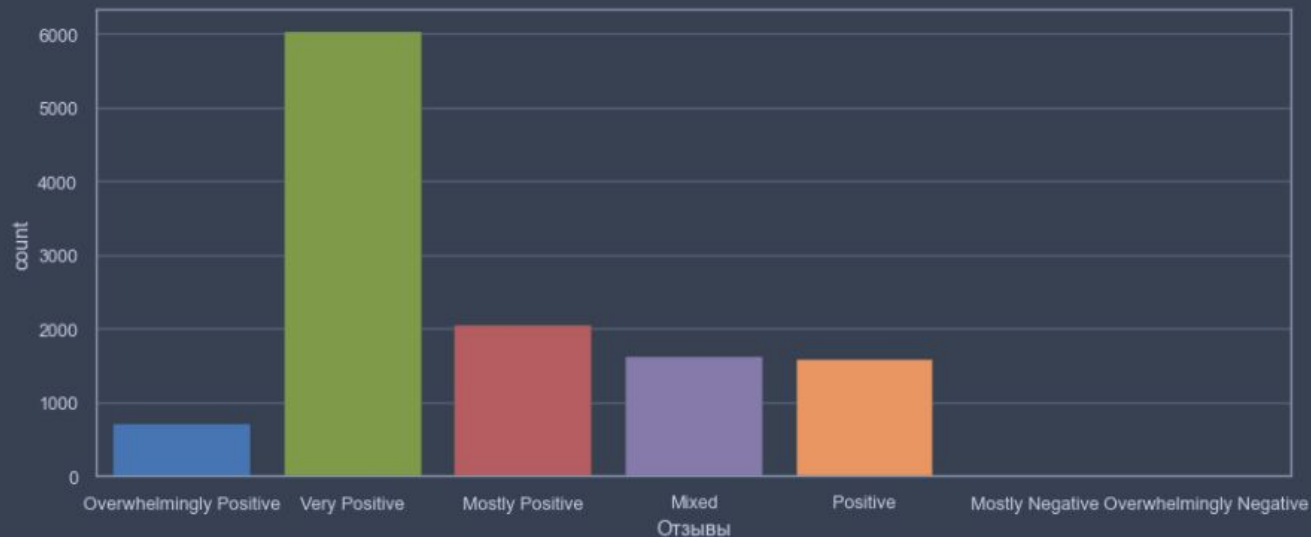
Самые дорогие игры

```
1 df.sort_values(by = 'Цена', ascending = False).head(10)
```

	Название	Цена	Метки	Тип	Дата выхода	Отзывы	Рейтинг	Число отзывов
10577	Mosaic: Game of Gods	99999.00	['Casual', 'Puzzle', 'God Game']	AAA	2016-10-31	Very Positive	94%	54
2260	Legend of Himari	10000.00	['Indie', 'Casual', 'Anime', 'Female Protagoni...]	indie	2018-03-19	Very Positive	89%	247
2509	The Orphan Dreams	9999.00	['Adventure', 'Casual', 'Indie', 'RPG']	indie	2016-04-27	Mostly Positive	78%	19
4292	Yandere School	6666.00	['Anime', 'Action', 'Indie', 'Sexual Content', ...]	indie	2017-09-01	Mixed	61%	273
769	RealFlight 9.5S	5098.32	['Simulation', 'Indie', 'Racing', 'Flight', 'VR']	indie	2021-12-07	Very Positive	80%	338
7674	A-Train: All Aboard! Tourism	4499.00	['Simulation', 'Trains', 'Immersive Sim', 'Cit...]	AAA	2021-12-08	Mostly Positive	70%	72
8475	Re:ZERO -Starting Life in Another World- The P...	4198.00	['Adventure', 'Anime', 'Visual Novel', 'Strate...]	AAA	2021-01-28	Very Positive	87%	377
6000	ELDEN RING	3999.00	['Souls-like', 'Relaxing', 'Dark Fantasy', 'RP...]	AAA	2022-02-24	Very Positive	89%	325,013
7161	Ys IX -Monstrum NOX-	3929.00	['Action', 'RPG', 'JRPG', 'Action RPG', '3D', ...]	AAA	2022-01-19	Very Positive	89%	69
9203	AFL Evolution 2	3699.00	['Sports', 'Simulation', 'Singleplayer', 'PvP'...]	AAA	2020-09-11	Very Positive	80%	122

Отзывы

```
1 sns.countplot(df['Отзывы'])
2 jtpplot.style(context='notebook', ticks=True, grid=True, fscale=1.2, figsize=(15,6), theme='oceans16')
```



```
1 df['Отзывы'].value_counts()
```

Very Positive	6027
Mostly Positive	2053
Mixed	1611
Positive	1582
Overwhelmingly Positive	715
Mostly Negative	4
Overwhelmingly Negative	2
Name: Отзывы, dtype: Int64	

Самый высокий рейтинг

```
1 df1 = df.copy()
2 df1[df1['Рейтинг'] == '100%']
```

	Название	Цена	Метки	Тип	Дата выхода	Отзывы	Рейтинг	Число отзывов
184	lb	299.0	[Female Protagonist, 2D, Horror, Cute, Psychol...	indie	2022-04-11	Very Positive	100%	197
1614	FIND ALL 2: Middle Ages	66.0	[Hidden Object, Casual, Hand-drawn, Colorful, ...	indie	2022-03-03	Positive	100%	36
1707	COSPLAY LOVE! : Enchanted princess	973.0	[Adventure, Visual Novel, Dating Sim, Choose Y...	indie	2022-03-30	Positive	100%	10
1763	Tentacular	699.0	[Casual, VR, Funny, Building, Puzzle, Action, ...	indie	2022-03-24	Positive	100%	10
1769	The Last Cube	435.0	[3D Platformer, Puzzle Platformer, Puzzle, Dif...	indie	2022-03-10	Very Positive	100%	55
...
11959	Corridor 7: Alien Invasion	259.0	[Retro, Aliens, FPS, Shooter, Pixel Graphics, ...	AAA	2021-08-06	Positive	100%	11
11970	Castle on the Coast	515.0	[3D Platformer, Collectathon, Retro, Cute, Old...	AAA	2021-12-02	Positive	100%	14
11978	Amazing Fix	259.0	[Early Access, Casual, Puzzle, Hidden Object, ...	AAA	2021-04-14	Positive	100%	20
11979	Mahjong Solitaire Refresh	435.0	[Casual, Puzzle]	AAA	2019-09-12	Positive	100%	10
11985	reiterate()	154.0	[Precision Platformer, Action, Fast-Paced, Pla...	AAA	2022-01-13	Positive	100%	32

394 rows × 8 columns

Общее распределение рейтинга

```
1 sns.distplot(df1['Рейтинг']).set_title('Распределение рейтинга')
```

```
Text(0.5, 1.0, 'Распределение рейтинга')
```



А теперь возьмём только те игры, у которых более 1000 ОТЗЫВОВ

	Название	Цена	Метки	Тип	Дата выхода	Отзывы	Рейтинг	Число отзывов
0	Phasmophobia	309.0	['Horror', 'Online Co-Op', 'Multiplayer', 'Psy...	indie	2020-09-18	Overwhelmingly Positive	97%	339441
1	Rust	3000.0	['Survival', 'Crafting', 'Multiplayer', 'Open ...	indie	2018-02-08	Very Positive	87%	620455
2	DayZ	1199.0	['Survival', 'Zombies', 'Multiplayer', 'Open W...	indie	2018-12-13	Mostly Positive	71%	249404
3	ARK: Survival Evolved	169.0	['Open World Survival Craft', 'Survival', 'Ope...	indie	2017-08-27	Very Positive	82%	410820
4	Risk of Rain 2	799.0	['Third-Person Shooter', 'Action Roguelike', '...	indie	2020-08-11	Overwhelmingly Positive	96%	130405
...
11883	Jesus Christ RPG Trilogy	0.0	['Illuminati', 'Memes', 'Psychological Horror'...	AAA	2016-03-18	Very Positive	85%	1477
11914	Bloodline Champions	0.0	['Free to Play', 'Action', 'MOBA', 'Multiplaye...	AAA	2011-10-28	Very Positive	85%	1646
11920	Disturbed	0.0	['Free to Play', 'Horror', 'Adventure', 'Visua...	AAA	2016-10-03	Very Positive	92%	1638
11934	War of the Three Kingdoms	0.0	['Sexual Content', 'Psychological Horror', 'Nu...	AAA	2021-12-17	Overwhelmingly Negative	10%	13705
11943	Audition Online	0.0	['Free to Play', 'Music', 'Rhythm', 'Massively...	AAA	2015-06-04	Mixed	61%	1472

2964 rows × 8 columns

Есть ли у них 100%?

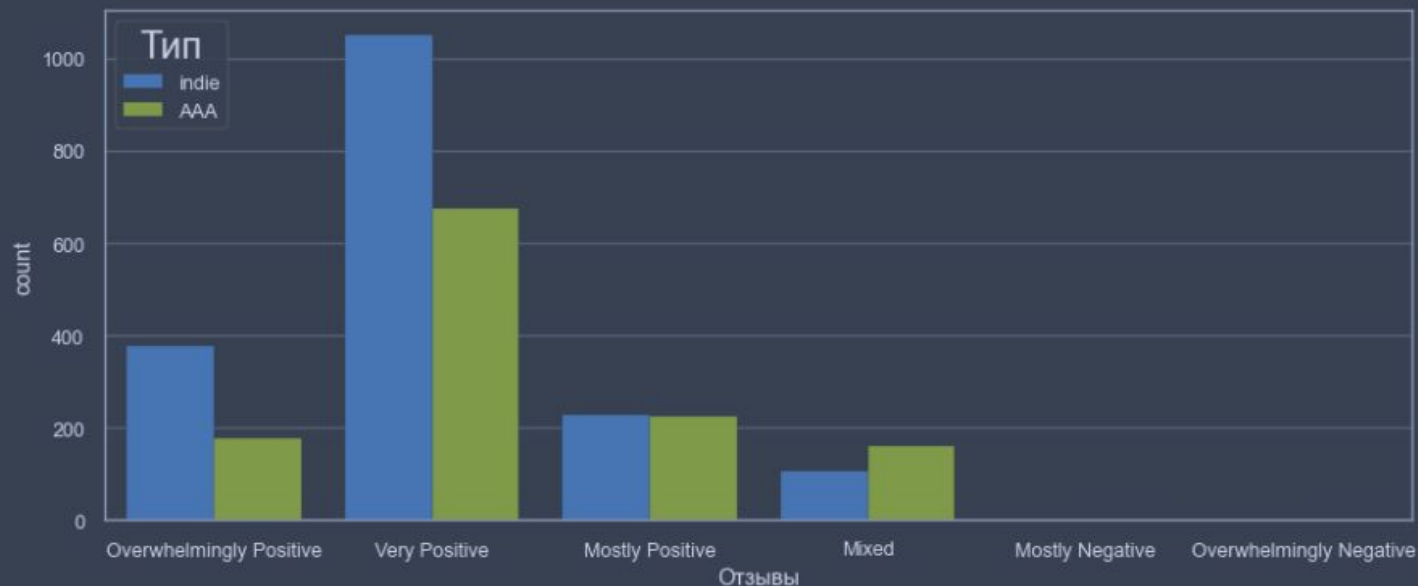
```
1 df_1000_review[df_1000_review['Рейтинг'] > 98]
```

	Название	Цена	Метки	Тип	Дата выхода	Отзывы	Рейтинг	Число отзывов
1052	A Short Hike	195.0	[Adventure, Exploration, Indie, Cute, Relaxing...	indie	2019-07-30	Overwhelmingly Positive	99.0	8473
4435	Aventura Copilului Albastru și Urât	61.0	[Adventure, Hentai, Sexual Content, 2D, Gore, ...	indie	2021-02-08	Overwhelmingly Positive	99.0	1815
6186	ULTRAKILL	435.0	[Early Access, FPS, Arena Shooter, Spectacle f...	AAA	2020-09-03	Overwhelmingly Positive	99.0	20380
6400	Senren * Banka	899.0	[Casual, Adventure, Visual Novel, Anime, Datin...	AAA	2020-02-14	Overwhelmingly Positive	99.0	7908
6664	Sucker for Love: First Date	259.0	[Visual Novel, Romance, Dating Sim, RPG, Horro...	AAA	2022-01-20	Overwhelmingly Positive	99.0	1142
6810	Maitetsu: Last Run!!	1459.0	[Sexual Content, Casual, Visual Novel, Anime, ...	AAA	2020-10-30	Overwhelmingly Positive	99.0	1586
11390	Project Kat - Paper Lily Prologue	0.0	[Anime, Psychological Horror, Pixel Graphics, ...	AAA	2021-10-16	Overwhelmingly Positive	99.0	2527

Зависимость между типом и отзывами

```
2 sns.countplot(x='Отзывы', hue='Тип', data = df_1000_review)
```

<AxesSubplot:xlabel='Отзывы', ylabel='count'>



Стереотип о пиксельной графике

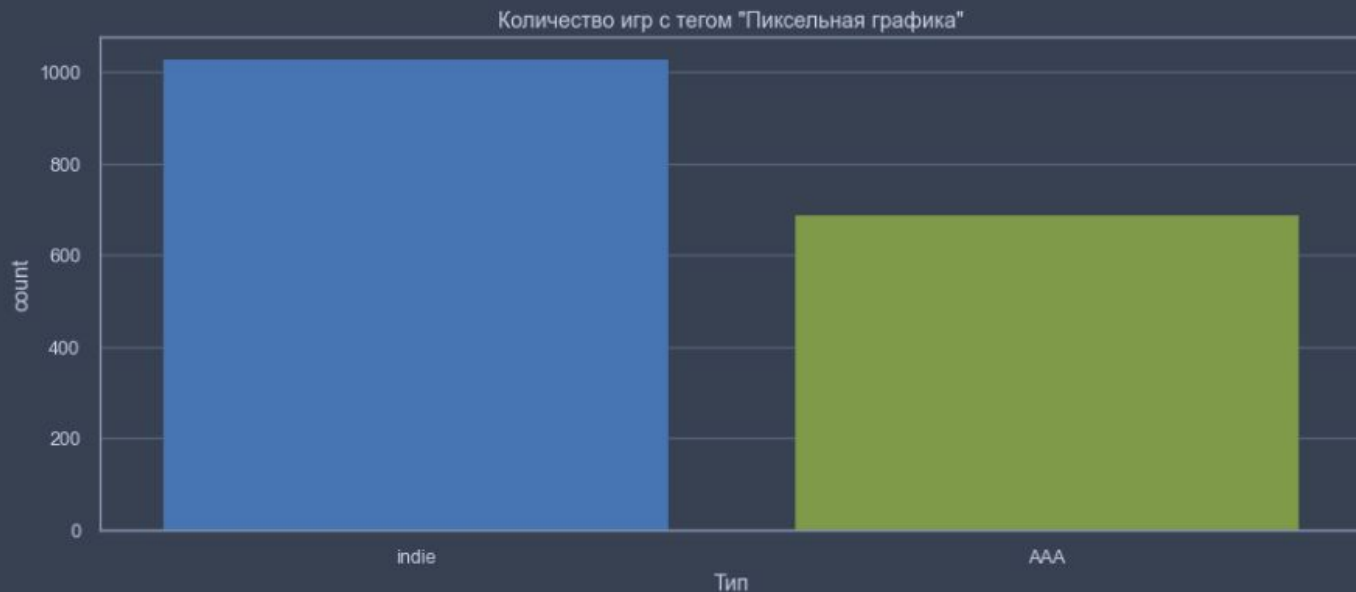
	Название	Цена	Метки	Тип	Дата выхода	Отзывы	Рейтинг	Число отзывов
6	Vampire Survivors	73.0	[Action Roguelike, Pixel Graphics, Bullet Hell...	indie	2021-12-17	Overwhelmingly Positive	98%	69,911
19	Stardew Valley	299.0	[Farming Sim, Life Sim, Pixel Graphics, RPG, M...	indie	2016-02-26	Overwhelmingly Positive	98%	381,999
23	The Binding of Isaac: Rebirth	449.0	[Action Roguelike, Roguelike, Indie, Replay Va...	indie	2014-11-04	Overwhelmingly Positive	97%	173,230
25	Terraria	610.0	[Open World Survival Craft, Sandbox, Survival,...	indie	2011-05-16	Overwhelmingly Positive	98%	745,746
26	Dead Cells	499.0	[Roguelite, Roguelike, Pixel Graphics, Difficu...	indie	2018-08-06	Overwhelmingly Positive	97%	89,192
...
11949	Black Widow: Recharged	809.0	[Action, Arcade, Shooter, Shoot 'Em Up, Twin S...	AAA	2021-10-28	Positive	91%	12
11959	Corridor 7: Alien Invasion	259.0	[Retro, Aliens, FPS, Shooter, Pixel Graphics, ...	AAA	2021-08-06	Positive	100%	11
11965	Neko Dungeon 喵酱迷城 喵酱迷城 ねこダンジョン	195.0	[JRPG, Sexual Content, Roguelite, Hidden Objec...	AAA	2018-08-17	Mixed	46%	30
11975	Super Grave Snatchers	154.0	[Villain Protagonist, Retro, Action RPG, Dark ...	AAA	2021-10-01	Positive	86%	36
11976	诛天记	0.0	[Casual, Action, RPG, 3D Fighter, Action RPG, ...	AAA	2021-07-15	Mixed	62%	159

1711 rows × 8 columns

Стереотип о пиксельной графике

```
1 sns.countplot(df3['Тип']).set_title('Количество игр с тегом "Пиксельная графика"')
```

```
Text(0.5, 1.0, 'Количество игр с тегом "Пиксельная графика"')
```



Итоги

- Средняя цена игры в инди-сегменте ниже средней цены AAA игр
- AAA игры составляют большую часть среди самых дорогих
- Игры с рейтингом 100% имеют, как правило, очень мало отзывов
- Среди всех игр с меткой «Пиксельная графика» большая часть будет принадлежать к инди



Извлечение данных из интернет-магазина Steam

Егорова Лада Владимировна,
444 группа

st068843@student.spbu.ru

14 апреля 2022 г.