

0 - Описание датасета

- Датасет `games_detailed_info.csv` собран [с сайта настолок](#).

Основные колонки данных: год(аж от до нашей эры до 2020), описание, количество игроков, возраст игроков, время игры и оценка от комьюнити по этим аспектам, тип и игровые механики, отношение пользователей. Также есть рейтинги, но в датасете они почти все пустые.

Все данные были типа ('O'), так что пришлось вручную все переправлять. (костыльно. очень.)

В целом по названиям колонок все понятно, кроме некоторых:

`averageweight` - сложность игры по оценкам пользователей

`wanting` - сколько человек хотят совершить трейд

`wishing` - находится в вишлисте

А по половине вообще непонятно зачем они. (На kaggle неправильные подписи к колонкам).

Большооой вопрос: Как создать настолку, которая всем понравится?

1 - Pandas

1. Пропуски

У многих колонок вообще 0 данных, поэтому дропаем их полностью и также убираем те, в которых более 50 процентов это NaN.

Кроме этого еще убираем колонки с неизвестными данными.

2. Выбросы

Данных, в которых может затаиться выброс не так много. Через scatterplot удобно находить их.

Удаляем какую-то игру, которой нужно 100 часов...

3 - 6 Работа с данными

1.3 Гипотеза: Для настолок с минимальным возрастом меньше 12 лет, среднее время игры будет короче

Ответ: Это правда, и почти в три раза!

Чем младше дети, тем меньше времени они способны фокусироваться на чем-то одном.

1.4 Вопрос: Настолки с каким количеством художников и количеством механик встречаются чаще всего?

Ответ: Одиночки и с одной механикой наиболее популярны.

Следующие строки в основном с двумя механиками, хоть они и сильно уступают играм, с одной механикой. Также можно подметить, что чаще всего рисуют настолки все-таки в одиночку.

1.5 Вопрос : Больше ли людей хотят настолку, если у нее длинное описание?

Ответ : Это правда, разница в символах три раза. Единственное, в конце остались настолки у которых 0 вхождений в вишлист, наверное надо было брать те, которых хотя бы один раз добавляли в вишлист.

1.6 Вопрос: В каком пятилетии вышло больше всего настолок?

Ответ: в 2015-2020 году

Также можно подметить, что в года рядом тоже довольно много настолок. Следовательно, настольные игры с каждым годом становятся все популярнее.

2 - Визуализация

2.1 Вопрос : Чем выше место в рейтинге, тем больше будет комментариев

Ответ: Действительно, чем популярнее игра тем больше людей оставляют комментарии. Но подъем кол-ва комментариев есть еще у самых последних в рейтинге. Возможно они уже отрицательные, а не

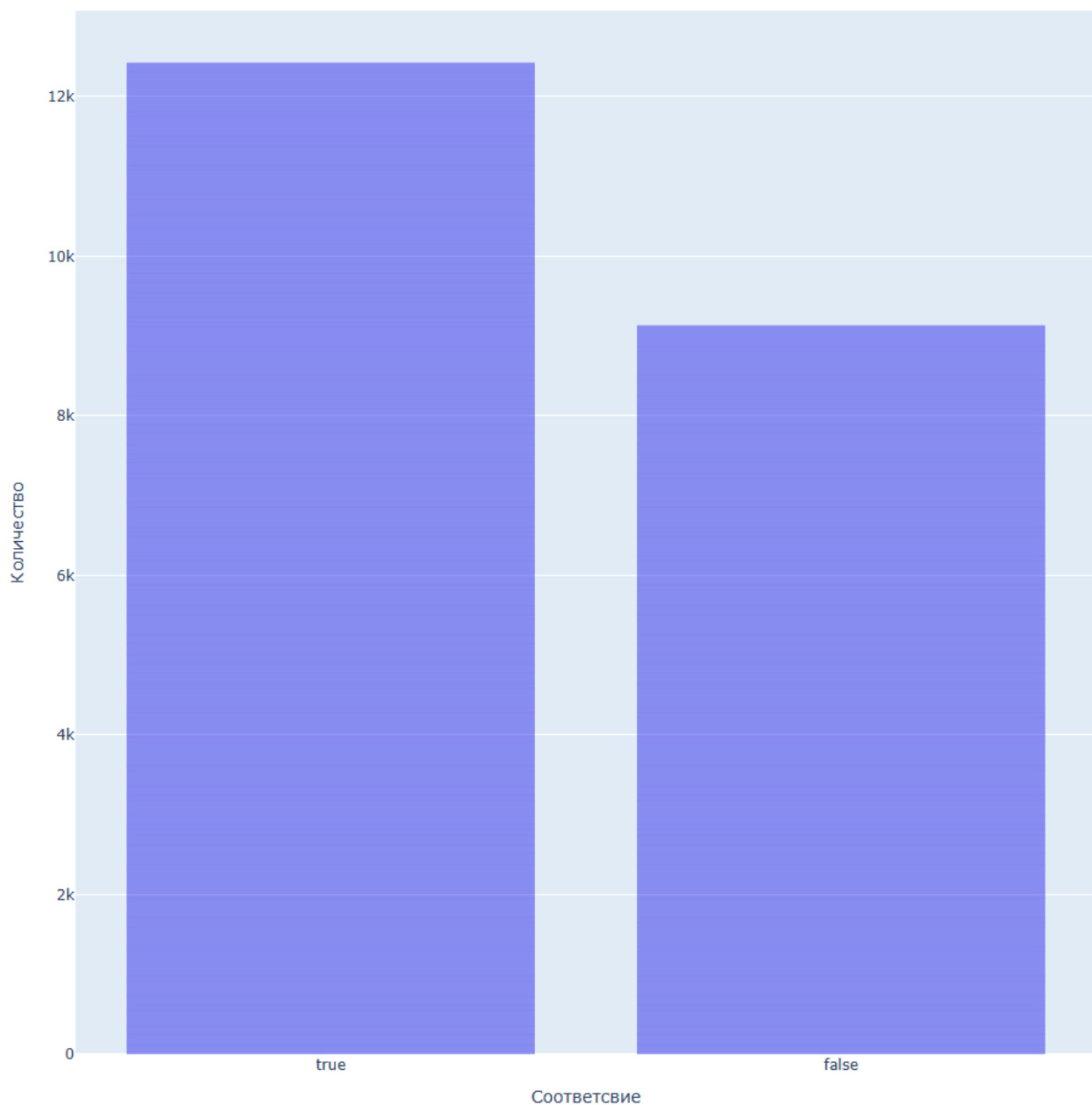
положительные(этого в датасете нет).



2.2 **Вопрос** На сколько соотносится среднее значение игроков в одной партии и реального оптимального количества из данных от комьюнити.

Ответ: БОльшая половина сходится, но все таки где-то 40 процентов не соответствует. Возможно стоит лучше ставить возрастные ограничения чтобы найти свою целевую аудиторию в будущем.

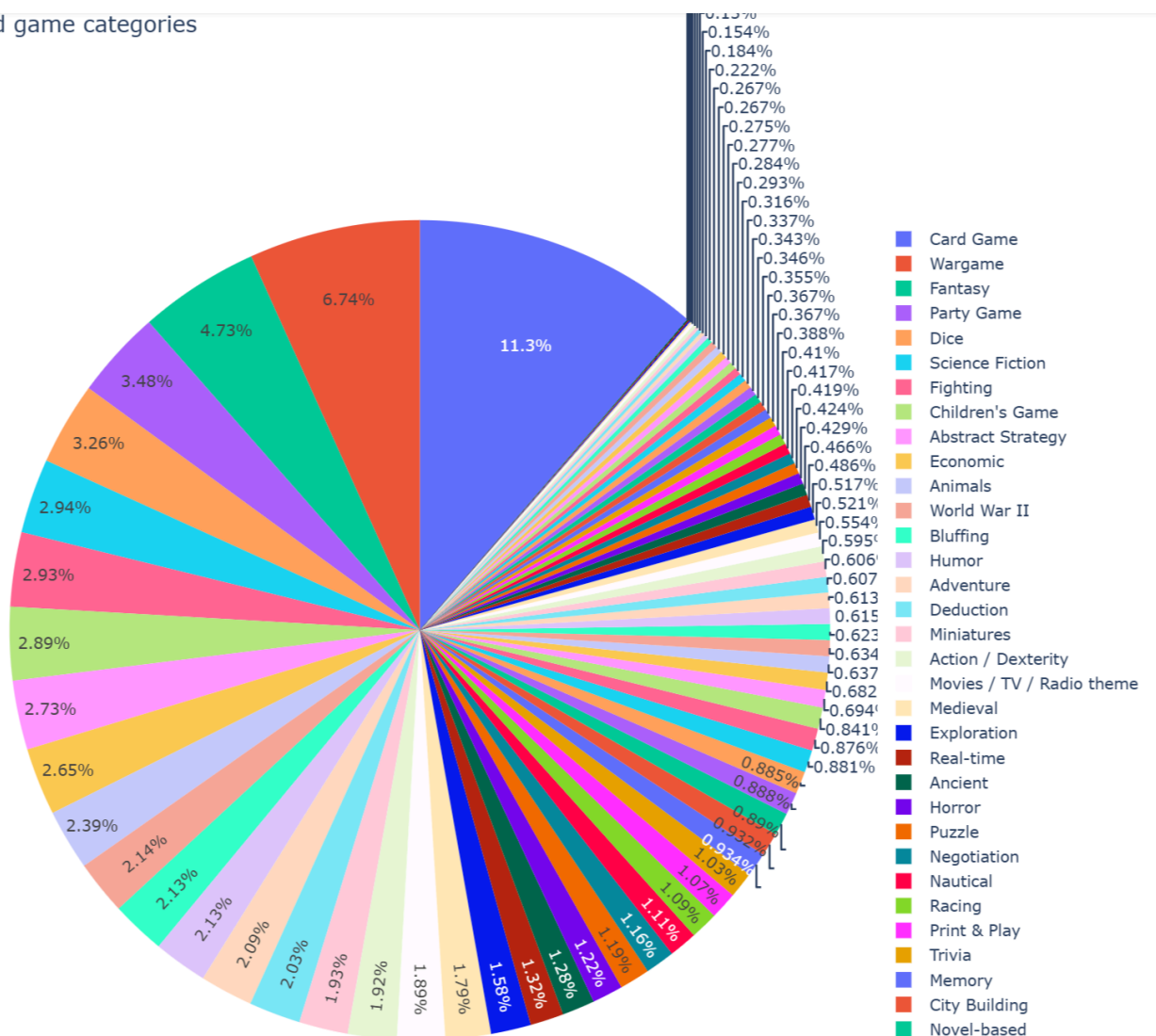
Соответствие ожидаемого идеального количества игроков с реальностью



2.3 Вопрос : Каковы доли различных игровых категорий? Какие чаще всего встречаются?

Ответ: Чаще всего встречаются карточные игры, что вполне логично. Стоит подметить, что здесь идет счет сколько раз вообще встречалась категория в датасете, ведь у одной игры может быть их несколько.

Board game categories

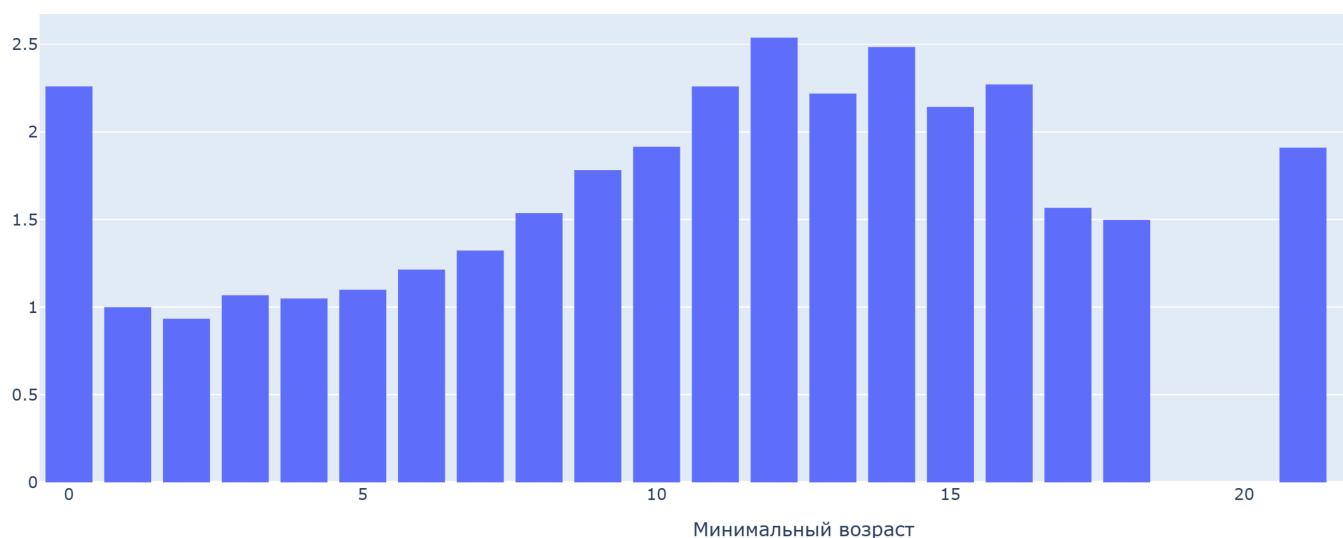


2.4 Гипотеза: Сложность игры зависит от самого младшего игрока

Ответ: Маленькие дети правят миром хох!

Гистограмма постепенно возрастает. Больше значение при "0" можно объяснить тем, что многие игры "family friendly" и предназначены для любого возраста, так что и получился такой результат.

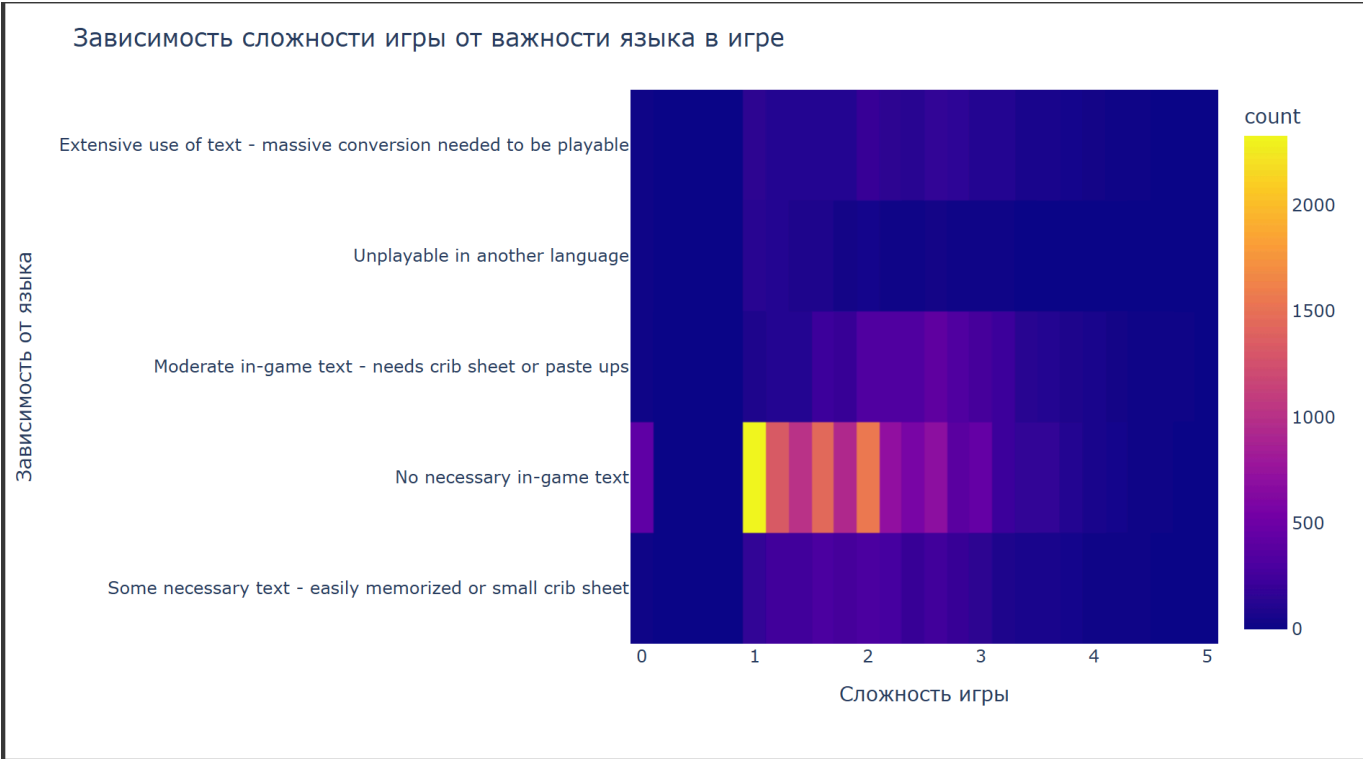
Средняя сложность игры от минимального возраста игрока



2.5 Гипотеза: Понимание языка (на картах и других реквизитах) делает игру легче.

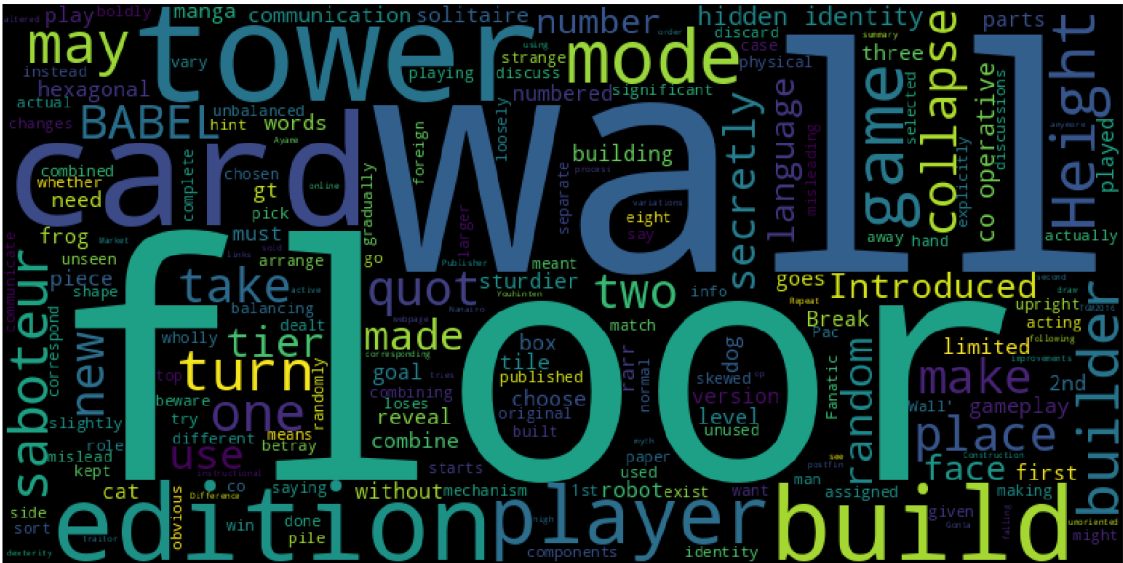
Ответ : Не самая удачная гипотеза.... Так получилось, что почти все игры имеют одинаковый рейтинг

сложности. Но посередине есть яркий кусочек, показывающий что игр без текста простого уровня очень много.



2.6 **Вопрос:** Какие слова встречаются у настолки с наибольшим описанием?
Тут целых 16к символов, а сама игра называется "Babel". Видимо разработчики, судя по облаку, решили записать туда всё правило игры... (а рейтинг у него 15017 из 2к)

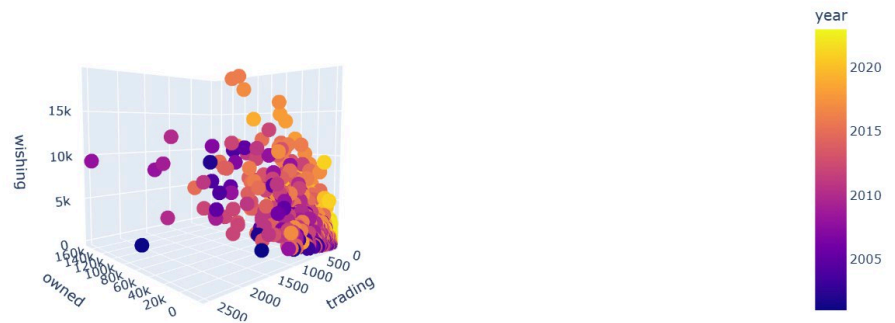
Longest description word cloud



2.7 **Вопрос** Как относится фидбэк (или же сколько людей предлагают трейд/хотят получить/уже имеют) настолки разных лет?(для 2000-2020)
Можно подметить, что чем новее игра, тем меньше людей ее имеет, и, соответственно предлагают трейд. Но в основном в вишлистах лежат довольно новые игры(оранжевые). Им уже 3-4 лет, они уже

набрали популярность и многие их хотят.

Настолки по годам и фидбэки аудитории



3 - Вывод

Обратно к большооому вопросу:

1. Хорошая тактика - попробовать пародировать те игры, которые уже популярны.
2. Еще одна тактика: написать правильную информацию для игроков, чтобы найти целевую аудиторию! Например семейная игра на 0+ рассчитана на широкий круг аудитории и скорее всего будет успешна
3. Попробовать сделать игру не очень сложной, но и не самой простой(уровня 1.8-2.5)
4. Написать хорошее описание. Слишком короткое не успеет заинтересовать пользователя, а слишком длинное как у 'Babel' , у которого не самый лучший рейтинг
5. Заметим, что все таки в основном в настолки играют люди до 25, так что надо выбрать подходящую тематику для них.
6. Писать просто и понятно на карточках настолок.