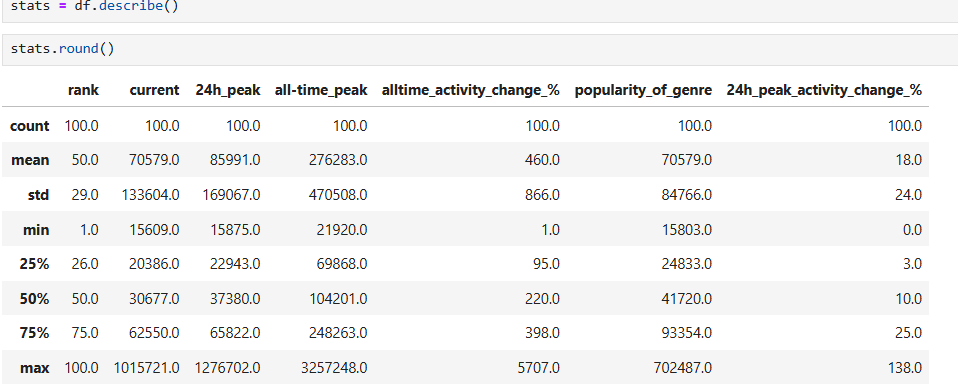
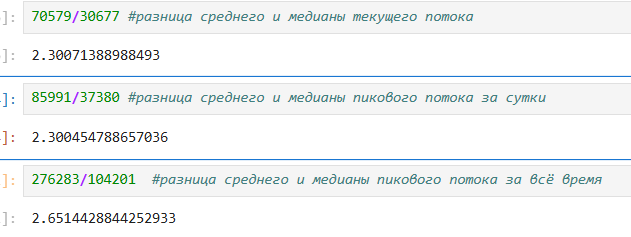
Отчет

Модуль Б «Анализ количественных данных»

**Анализ основных свойств данных**

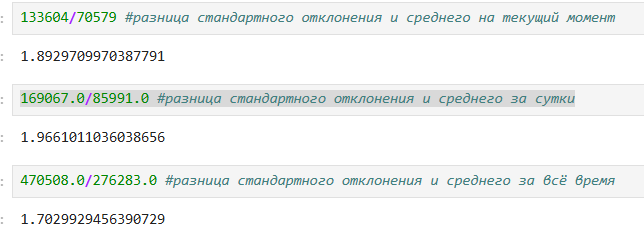


С помощью метода библиотеки pandas формирую таблицу со статистическими показателями атрибутов, где отображено общее количество параметров в каждом столбце (count); среднее значение (mean); стандартное отклонение (std); минимальное (min) и максимальные значения (max); квантили 25%, 50% (медиана), 75%. Для того, чтобы проанализировать состояние потоков по играм в разные периоды времени.



Среднее значение (mean) текущего количества онлайн игроков в 2 раза больше медианы (50%), что свидетельствует о наличии выбросов в данных. Также как и пик за сутки. Что касаемо пика за всё время, показатели отличаются более чем в 2,5 раз.

Можно предположить, что большинство игр имеют небольшой онлайн, но есть сверхпопулярные игры.



Стандартное отклонение (std) параметров current, 24h\_peak, all-time\_peak , больше среднего почти в 2 раза, означает, что среднее не отображает типичное значение, а медиана будет ближе к "реальному" среднему, значит точно есть игры с аномально большим онлайн-потоком.







Данные разнообразны, ведь самая популярная (max) игра по онлайн-пользователям значительно больше наименее (min) популярной.



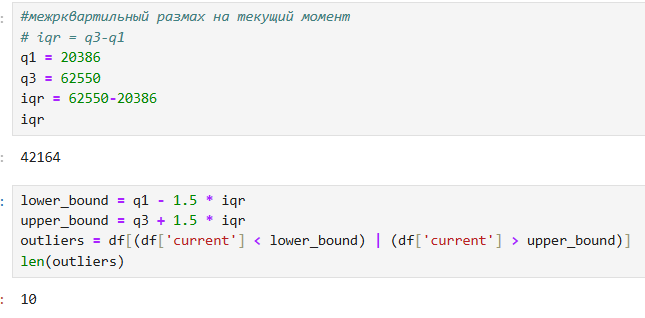


75-й квантиль говорит о том, что 75% игр имеют онлайн ниже 62 550 в текущий момент; ниже 65 822 за сутки; ниже 248 263 за всё время.

50-й квантиль говорит о том, что половина игр имеет онлайн-поток ниже 30 677 на текущий момент; ниже 37 380 за сутки; ниже 104 201 за всё время.

25-й квантиль говорит о том, что - 25% игр имеют онлайн меньше 20 386 на текущий момент; меньше 22 943 за сутки; меньше 69 868 за всё время.

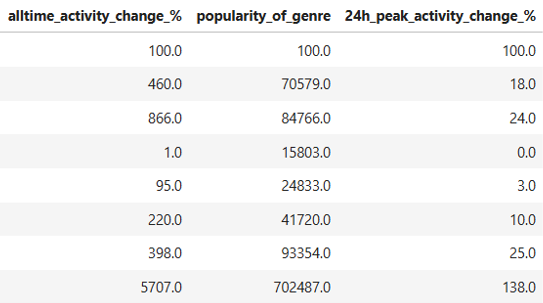
Значит можно сделать вывод о том, что данные ассиметричны, а меньшая часть имеет большой онлайн-поток.





Чтобы оценить разброс данных в выборке и подтвердить гипотезу о наличии сверхпопулярных игр, проведен расчет межквартильного размаха.

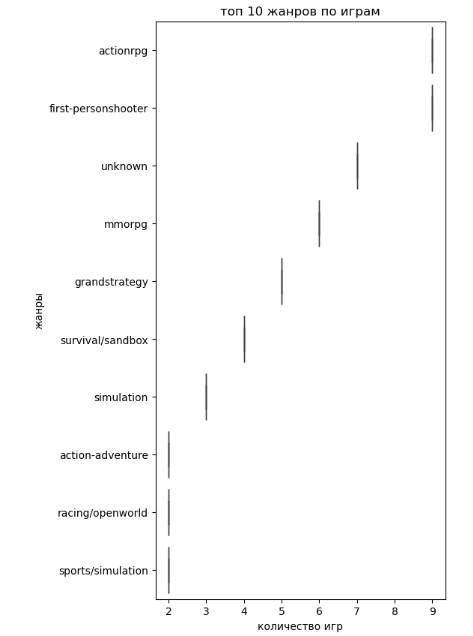
Межквартильный размах показывает, что средняя половина находится в диапазоне 42 164 на текущий момент; 42 879 за сутки; 178 395 за всё время. С помощью разбиения на интервалы выявлено от 10-12 выбросов, что и предполагалось раннее.



Пиковое изменение онлайнов за сутки Среднее процентное изменение = 18%. Медиана = 10% . Значения близки друг к другу. Стандартное отклонение = 24% Разница между мин и макс говорит о выбросах, значит есть те игры, где активность никак не изменилась и наоборот, где рост был занчительным.

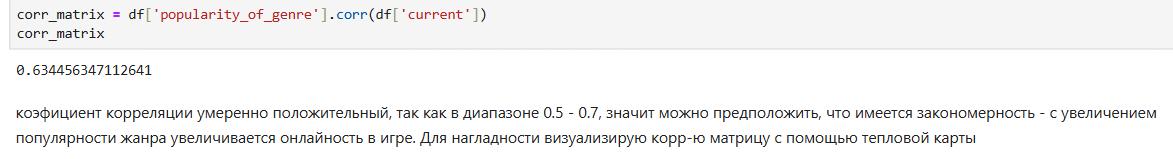
Минимальное = 0% это значит, что есть игры у которых не было ни роста ни падения активности. Максимальное = 138% это означает, что есть игра, где онлайн активность увеличилась в 2 раза.

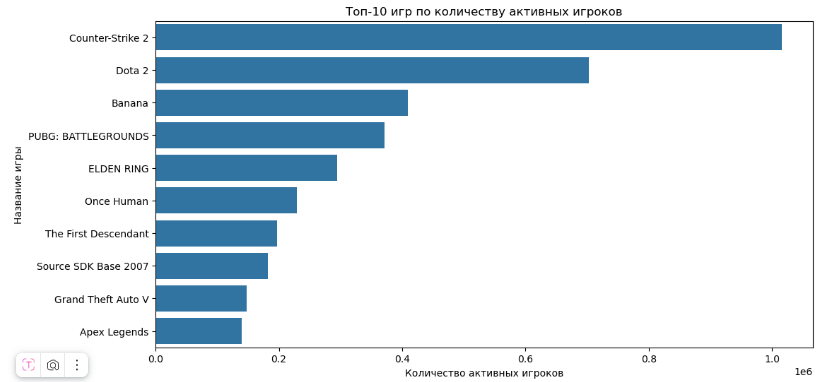
Межквантильный размах 25-3 = 22% Ассимметричное распределение, так как диапазон между 2 и 3 квантилями больше 1 квантиля в 2 раза



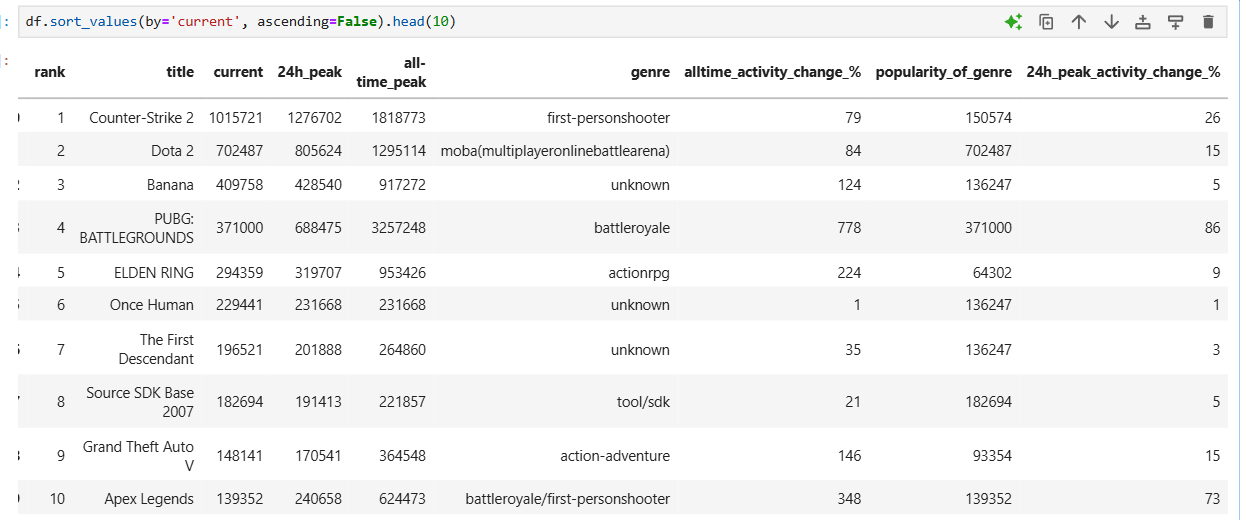
На таблице видно топ 10 жанров, наиболее популярные – actionrpg и first-personshooter, так как наиболее часто встречаются среди игр. Можно предположить, что онлайн-потоки, а соответственно популярность игры зависят от её жанра. Проверить можно выборкой топ 10 игр.

Проверка с помощью расчёта корреляции



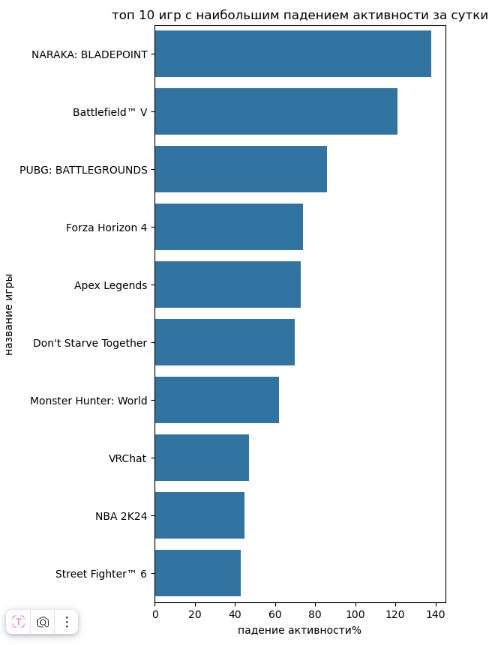


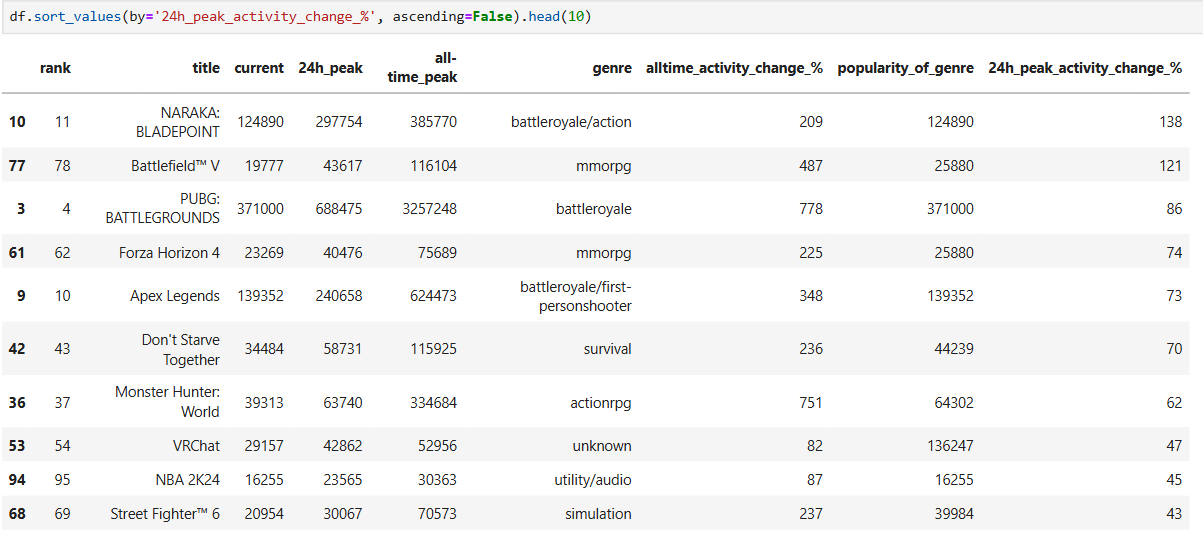
На текущий период наиболее популярная игра – Counter-Strike 2 жанра first-personshooter



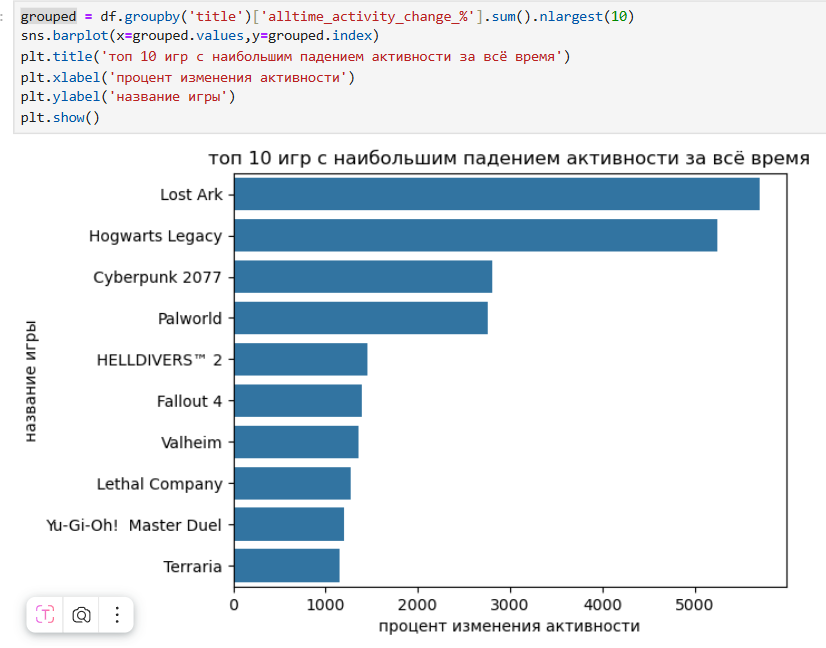
Сводная таблица по топ 10 игр на текущий момент, видно, что в них входит 2 игры first-personshooter.

Падение активности у Counter-Strike 2, Dota-2, Once Human, The First Descendant, Source SDK Base 2007 не значительно и показатели current, 24h\_peak, all-time\_peak изменяются симметрично



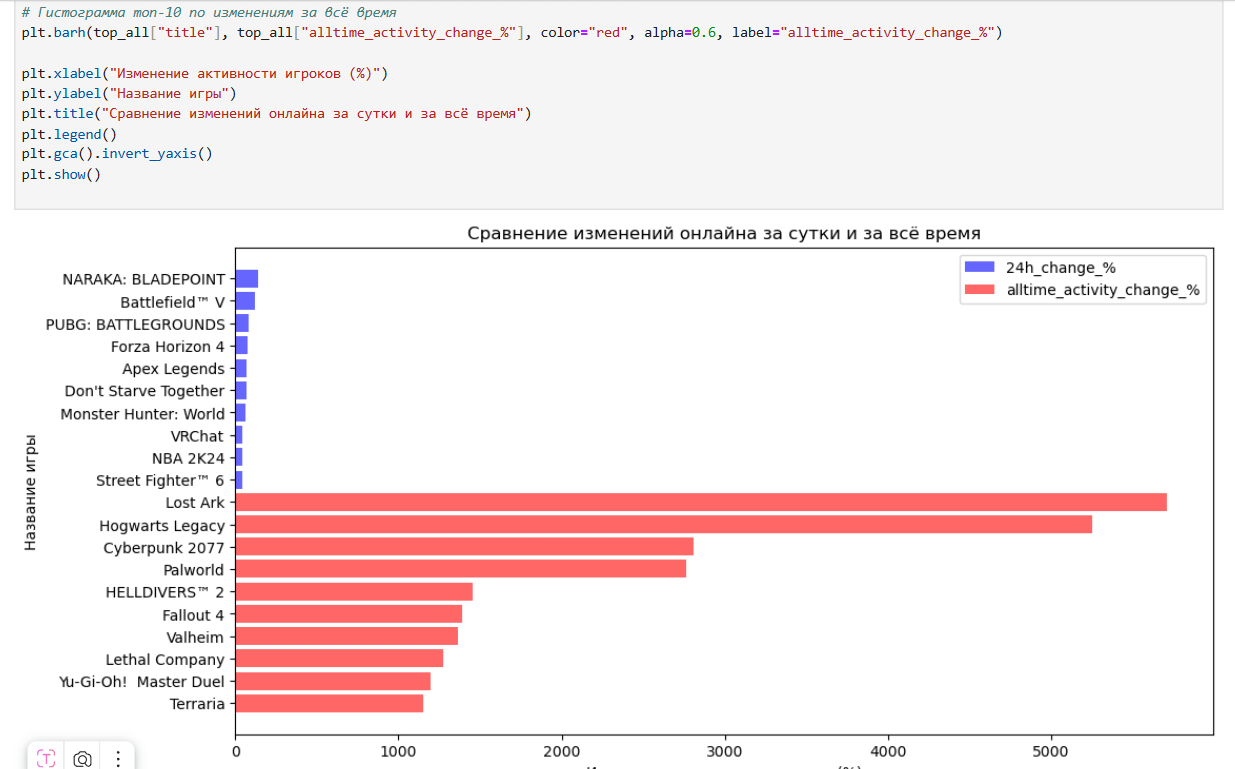


Сводная таблица по топ 10 играм с наибольшим падением активности за сутки, среди них есть игры из топа популярных жанров.





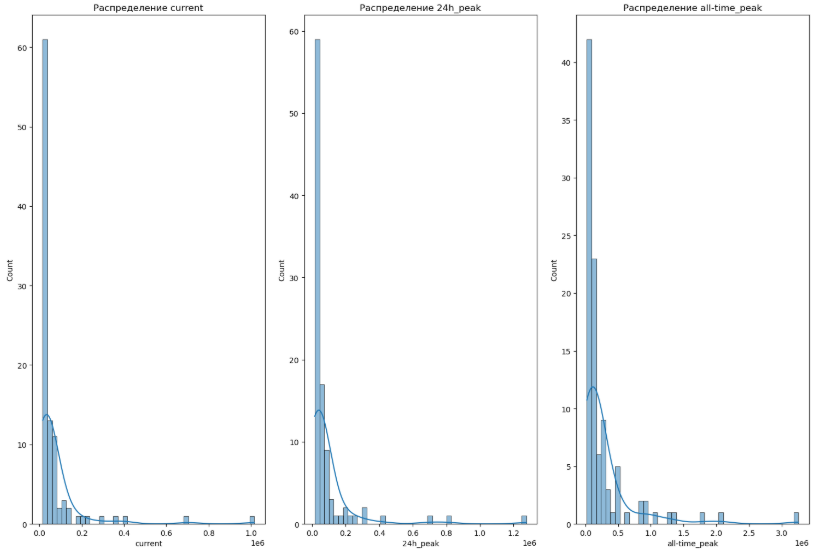
Сводная таблица топ 10 игр по падению активности за всё время. Исходя из показателей, произошёл обвал онлайн-потоков.

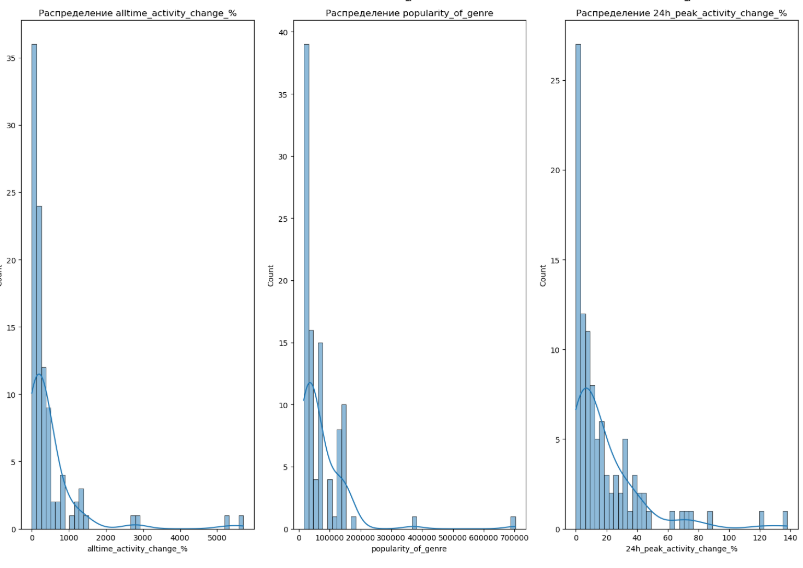


Построила диаграмму исходя из двух сводных таблиц, чтобы проверить совпадают ли игры в двух топах. Игры не совпадают.

Исходя из большого падения, можно предположить, что игры из топ 10 за всё время пережили свой пик популярности и сохраняют остаточную популярность.

Игры из топ 10 за сутки могли быть временным трендом.

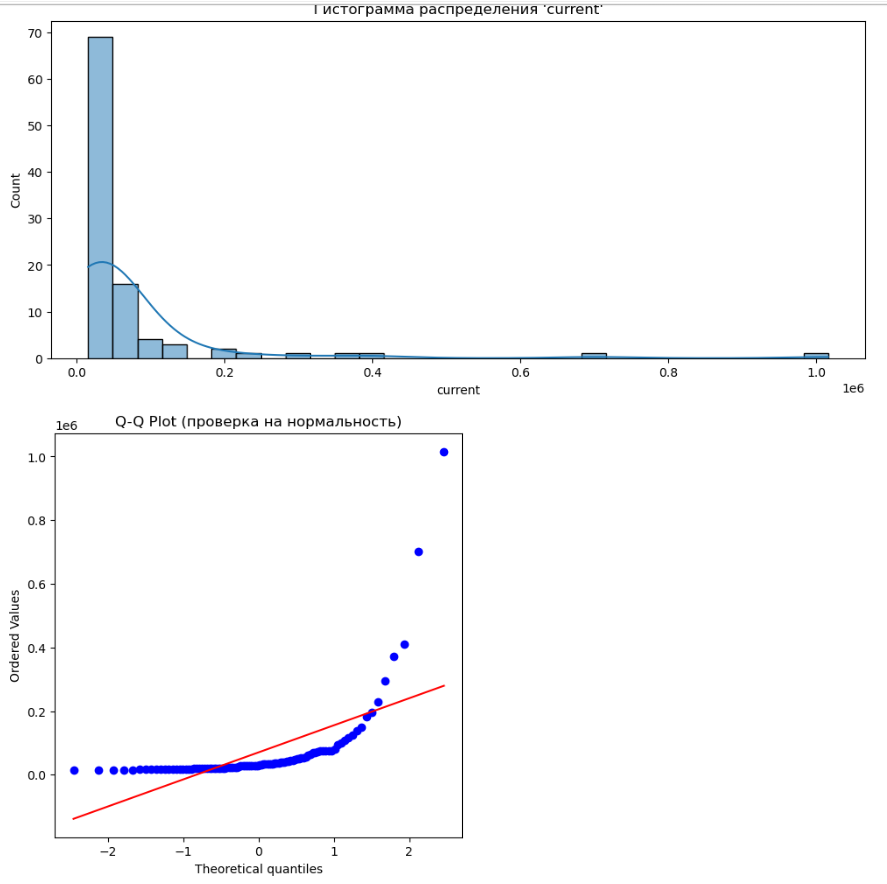




Определить тенденцию распределения активных пользователей в игровых продуктах можно на основе жанра, так как в топ 10 по активности входят игры одинакового жанра, по всем трем категориям онлайнов (current, 24h\_peak, all-time\_peak). Можно предположить, что в следующем году

**Проверка нормальности распределения**:

На втором графике QQ (квантиль-квантиль) данные (точки) не идут по диагонали, что свидетельствует о не нормальном распределении данных.



Вывод:

Можно сделать вывод о том, что

Наименее популярная игра на текущий момент- Victoria 3. Наиболее популярная игра на текущий момент - Counter-Strike 2 (жанр first-personshooter.).

Наименее популярная игра за сутки - Supermarket Simulator (vampire/openworld). Наиболее популярная игра за сутки все также - Counter-Strike 2 (жанр first-personshooter.).

Наименее популярная игра за сутки – Soundpad(simulation/tycoon). Наиболее популярная игра за всё время - PUBG: BATTLEGROUNDS (жанр battleroyale) (в топе текущей активности и за сутки держалась на 3-4 месте).

Предположительно, активность у игр жанров battleroyale и first-personshooter, и actionrpg может возрасти активность, так как они держались в топах 10 по наибольшей онлайн-активности, также игры №1 в топах соответствовали этим жанрам.

Игры жанров vampire/openworld и simulation/tycoon, grandstrategy наихудшие по всем категориям, активность стабильно низкая и не изменяется.