Отчет по тестовому заданию на позицию

«Младший продуктовый аналитик»

Демин Даниил Игоревич

Задача аналитика: Подтвердить или опровергнуть высказанные гипотезы на основании расчетов. Презентовать результаты анализа, подтвердив их таблицами, возможно, графиками.

Для корректного анализа из объединенного дата фрейма с данными о гильдиях и энергией гильдии были удалены строки, где количество участников = 0. Таких строк было 131.

**Шаг 1:** Определить критерии активности гильдий на основании файлов guild\_data.csv, energy.csv. Самостоятельно выбрать метод:

Мною был выбран метод кластеризации методом ближайших соседей с учетом экспертного мнения, и так же, экспериментально, метод кластеризации K-Means. На мой взгляд, метод ближайших соседей будет лучше, нежели экспертно на основании средних и квартилей различных параметров по следующим причинам:

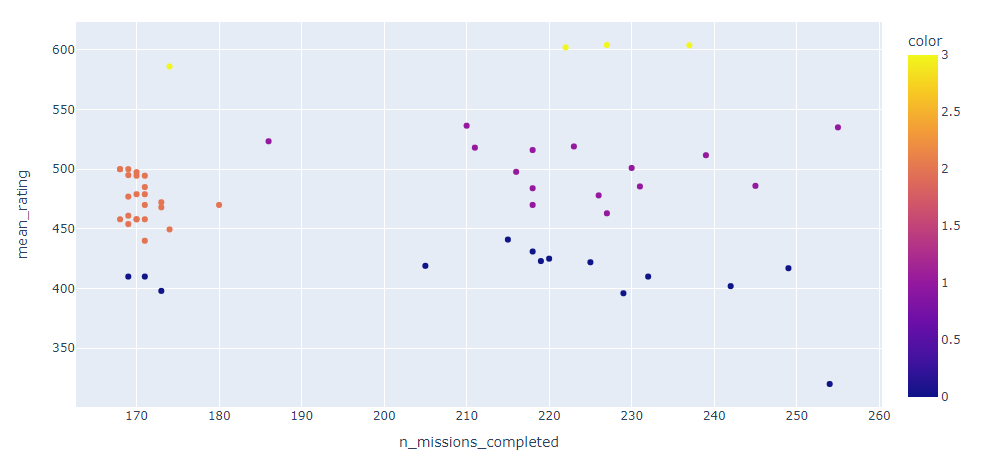
* Возможность учитывать экспертное мнение для того, чтобы предварительно разделить выборки по близким и важным параметрам, что может быть важным для правильного и более предсказуемого формирования кластеров
* Позволяет учитывать несколько параметров и особенности каждого, что было полезно, т.к. данные имели различные характеристики
* Метод менее чувствителен к выбросам в данных, чем тот же K-Means, что позволяет получить более стабильные результаты
* Большим плюсом так же является то, что метод менее сложен, т.к. он просто находит ближайшие объекты и объединяет их в кластеры

Во-первых, из всей выборки я отобрал гильдии, где было более 0 участников (разница между n\_added\_characters и n\_removed\_characters).

Так же я провел кластеризацию по нескольким параметрам 3 раза для того, чтобы оценить, насколько будут отличаться результаты в зависимости от созданных кластеров по разным параметрам.

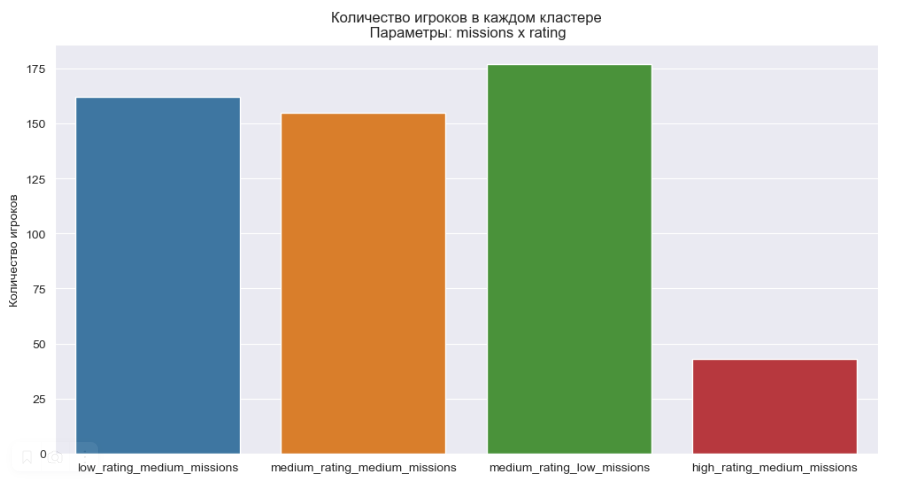
Критерии, которые были мною, определены для различных групп:

1. Первая выборка: n\_missions\_completed и mean\_rating для того, чтобы оценить исключительно соревновательные гильдии, насколько активно они принимают участие в войнах гильдии и интересуются другими внутри игровыми активностями. Было сформировано 4 кластера
   * low\_rating\_medium\_missions - Низкий рейтинг, среднее количество выполненных миссий
     + Средний рейтинг < 450
     + Количество выполненных миссий < 260
   * medium\_rating\_medium\_missions - Средний рейтинг, среднее количество выполненных миссий
     + Средний рейтинг < 550
     + Количество выполненных миссий < 260
   * medium\_rating\_low\_missions - Средний рейтинг, малое количество выполненных миссий
     + Средний рейтинг <= 500
     + Количество выполненных миссий <= 180
   * high\_rating\_medium\_missions - Высокий рейтинг, среднее количество выполненных миссий
     + Средний рейтинг > 550
     + Количество выполненных миссий > 170

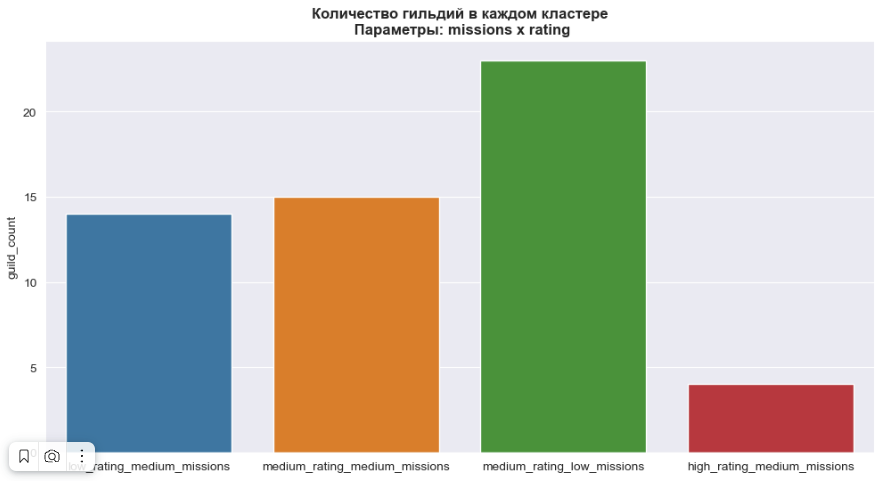


Распределение гильдий по среднему рейтингу и количеству выполненных миссий

Данная визуализация показывает нам, что у нас имеются разнообразные гильдии, которые, например, преимущественно участвуют в войнах гильдий, но мало участвуют в других активностях или практически не участвуют в рейтинге и занимаются лишь другими активностями



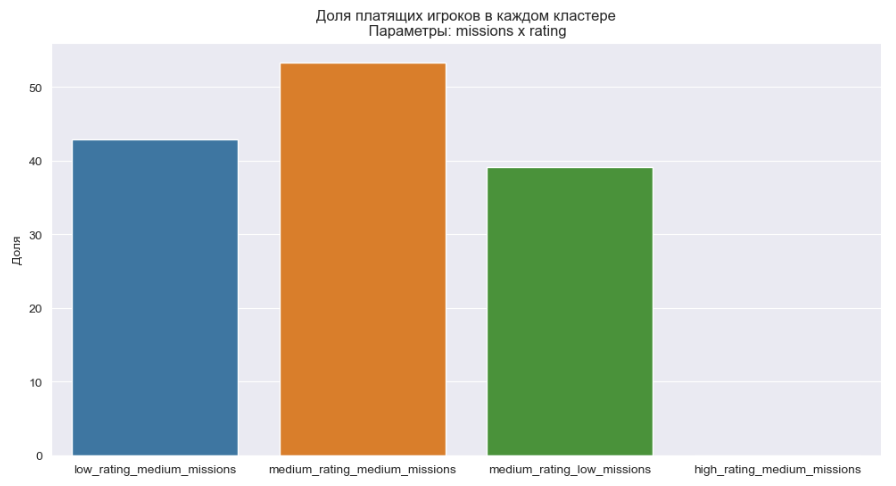
Данная же визуализация показывает нам наглядно, сколько игроков находится в том или ином кластере и можно предположить, что большая часть игроков стараются принимать активное участие в войнах гильдий, но на другие активности гильдии обращают меньше внимания



Данная визуализация показывает наглядно, сколько гильдий, в каком кластере расположилось.

**Гипотеза:** *большинство создаваемых гильдий не активны*.

**Ответ:** Из гильдий, которые принимают активное участие в войнах гильдий и других внутри игровых активностях, большинство активны и принимают активное участие именно в войнах гильдии, но уделяют меньше внимания другим активностям



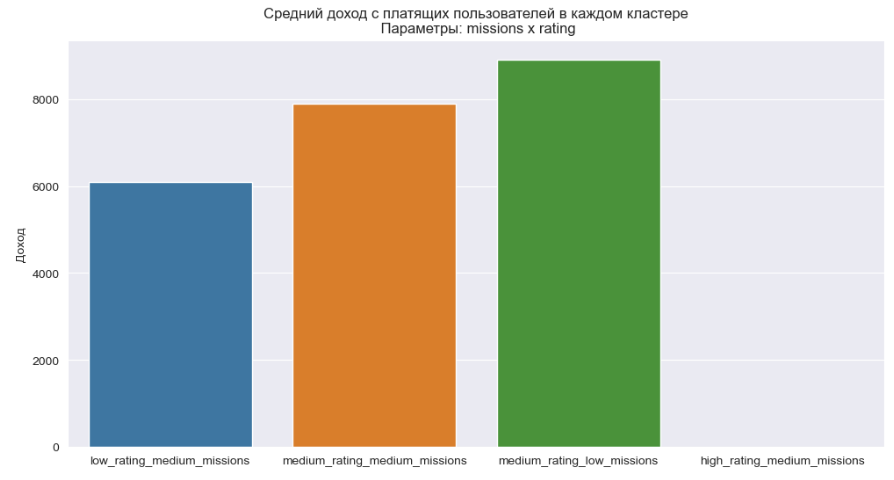
Данная визуализация показывает нам долю платящих игроков в каждом кластере. Что интересно, в данной категории мы рассматриваем лишь те гильдии, которые принимают активное участие в войнах гильдий, занимают места, получают рейтинг и гильдии, занимающие самый высокий рейтинг, не имеют платящих игроков, в то время как гильдии низкого рейтинга имеют долю платящих игроков свыше 40%

|  |  |
| --- | --- |
| **Кластер** | **share\_payers** |
| **low\_rating\_medium\_missions** | 42.86 |
| **medium\_rating\_medium\_missions** | 53.33 |
| **medium\_rating\_low\_missions** | 39.13 |
| **high\_rating\_medium\_missions** | 0.00 |

Таблица расчета доли платящих игроков для каждого кластера

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **character\_id** | **n\_players** | **n\_missions\_completed** | **mean\_rating** | **Clusters\_k3** | **revenue** |
| 13161077 | 20 | 227.0 | 604.00 | 3 | NaN |
| 19881930 | 4 | 237.0 | 603.75 | 3 | NaN |
| 22655914 | 18 | 222.0 | 602.00 | 3 | NaN |
| 21140835 | 1 | 174.0 | 586.00 | 3 | NaN |

Таблица, показывающая кластер гильдий с самым высоким рейтингом и средним количеством выполненных миссий



Данная визуализация показывает нам метрику ARPPU – средний доход с платящего игрока.

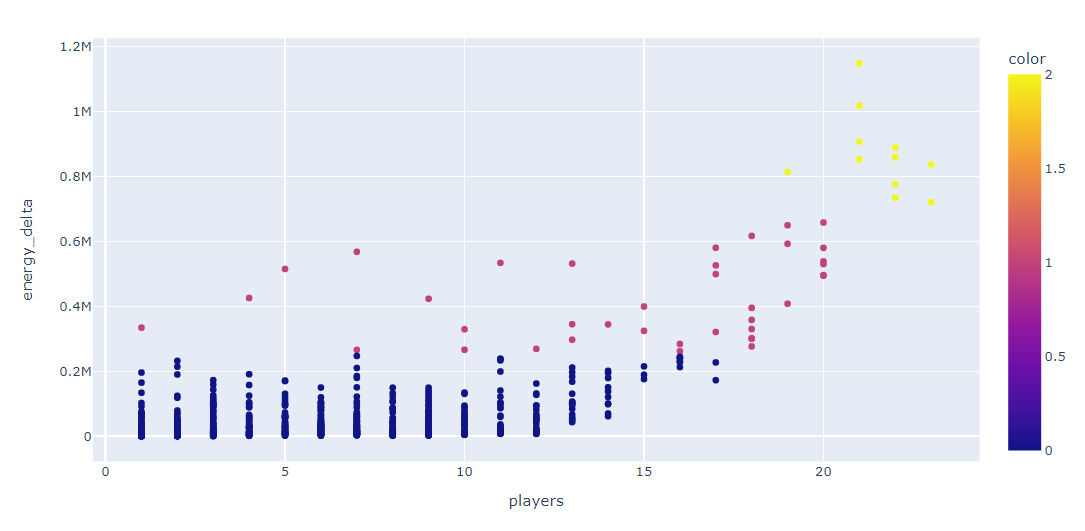
Интересно отметить, что, несмотря на то, что в гильдиях со средним рейтингом и средним количеством выполненных миссий доля платящих значительно выше, чем в гильдиях со средним рейтингом и низким количеством выполненных миссий, ARPPU у вторых выше.

|  |  |
| --- | --- |
| **Кластер** | **ARPPU** |
| **low\_rating\_medium\_missions** | 6090.21 |
| **medium\_rating\_medium\_missions** | 7887.71 |
| **medium\_rating\_low\_missions** | 8909.60 |
| **high\_rating\_medium\_missions** | NaN |

**Гипотеза***: игроки, покупающие премиум товары, создают более активные гильдии*, поскольку, инвестируя в игру, пользователь больше ценит вложенные усилия.

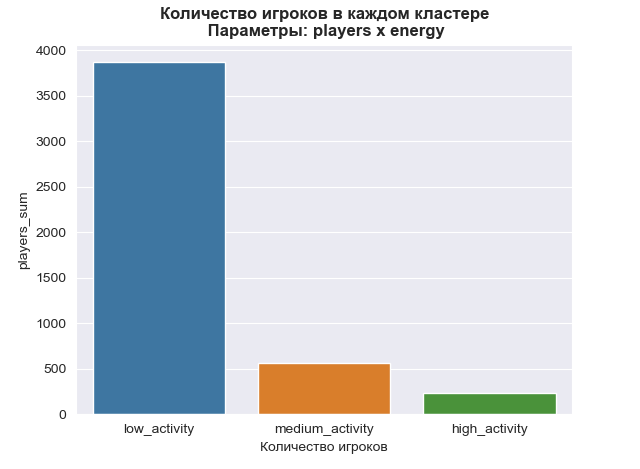
**Ответ:** В данном случае, рассматривая только те гильдии, которые принимают активное участие в войнах гильдий, самые активные гильдии вообще не имеют платящих игроков. Но наибольший средний доход за платящего пользователя находится в гильдиях, которые имеют средний рейтинг и низкое/среднее количество выполненных миссий.

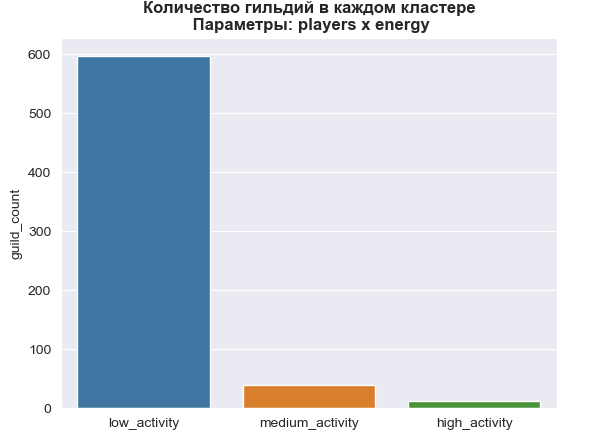
1. Вторая выборка: n\_players и energy\_delta (n\_players – вычисляемый столбец, сколько находится игроков в каждой гильдии), для того, чтобы оценить, насколько количество игроков влияет на получаемое количество энергии для гильдии
   * low\_activity - Низкая активность
     + Количество игроков < 17
     + Энергия гильдии < 260к
   * medium\_activity - Средняя активность
     + Количество игроков <= 20
     + Энергия гильдии < 700к
   * high\_activity - Высокая активность
     + Количество игроков > 19
     + Энергия гильдии > 700к



Распределение гильдий по количеству игроков и энергии гильдии

Данная визуализация показывает нам, что у нас имеется большинство гильдий, где до 15 участников и количество энергии гильдии мало, что свидетельствует о малой активности





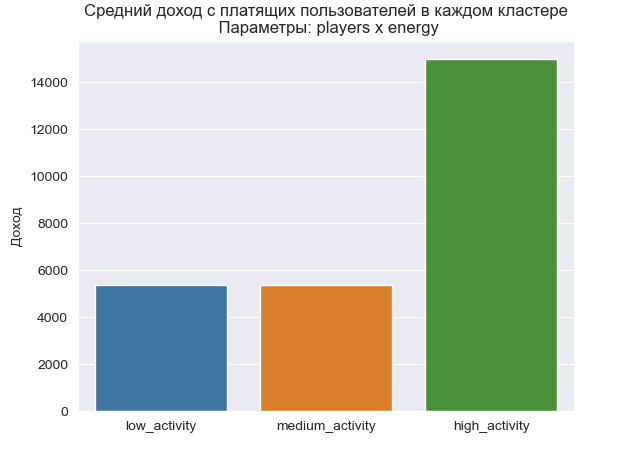
Преимущественное количество игроков и гильдий находятся именно в кластере с низкой активностью

**Гипотеза:** *большинство создаваемых гильдий не активны*.

**Ответ:** В данной категории большинство гильдий не активны или имеют низкую активность



|  |  |
| --- | --- |
| **Кластер** | **share\_payers** |
| **low\_activity** | 35.51 |
| **medium\_activity** | 50.00 |
| **high\_activity** | 54.55 |



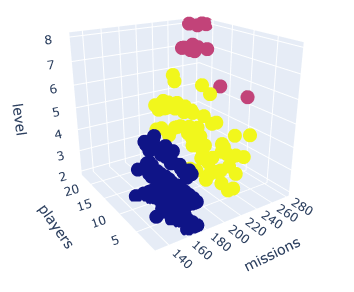
|  |  |
| --- | --- |
| **Кластер** | **ARPPU** |
| **low\_activity** | 5357.48 |
| **medium\_activity** | 5367.58 |
| **high\_activity** | 14999.39 |

Данная визуализация показывает средний доход с платящих игроков в каждом кластере. Учитывая предыдущие графики, можно прийти к заключению, что, несмотря на огромное количество игроков, гильдий и 35% платящих пользователей, кластеры низкой и средней активности даже в совокупности не достигают среднего дохода с кластера высокой активности, хотя там находится меньше всего игроков и гильдий

**Гипотеза***: игроки, покупающие премиум товары, создают более активные гильдии*, поскольку, инвестируя в игру, пользователь больше ценит вложенные усилия.

**Ответ:** В данном случае, рассматривая все гильдии, имеющие не пустыми рассмотренные параметры, платящие игроки создают самые активные гильдии.

1. Третья выборка: n\_missions\_completed и n\_players и level - для того, чтобы оценить, какого уровня чаще всего достигают гильдии, сколько в них игроков и насколько им интересно принимать участие во внутри игровых активностях. Не пропадает ли интерес с уровнем гильдии
   * low\_activity – Слабая активность/неактивность
     + Количество выполненных миссий < 190
     + Количество игроков <= 15
     + Уровень гильдии <= 4
   * medium\_activity – Средняя активность
     + Количество выполненных миссий <= 235
     + Количество игроков <= 20
     + Уровень гильдии <= 6
   * high\_activity – Высокая активность
     + Количество выполненных миссий > 235
     + Количество игроков >= 5,
     + Уровень гильдии >= 6



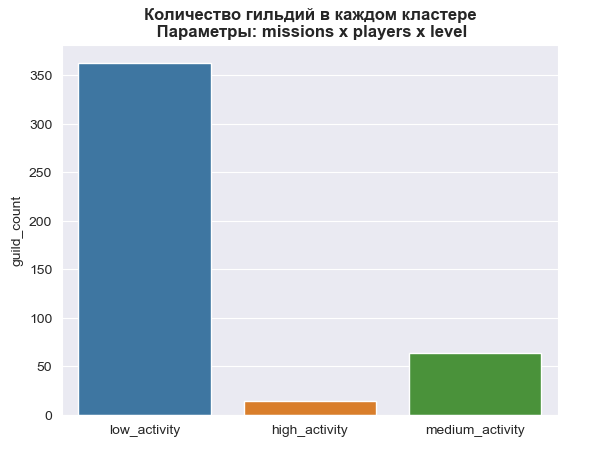
Распределение гильдий по количеству игроков, выполненным миссиям и уровню гильдии

Данная визуализация иллюстрирует нам распределение игроков по 3 параметрам:

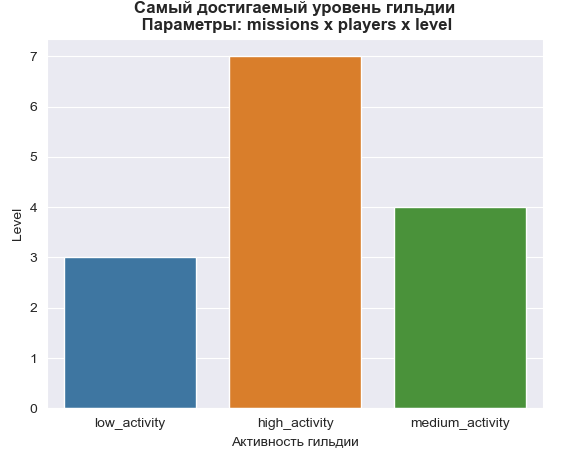
* количество завершенных миссий
* количество игроков
* уровень гильдии

Можно заметить, что большинство гильдий достигают 4 уровня гильдии, имеют до 15 игроков и выполняют 160-220 миссий





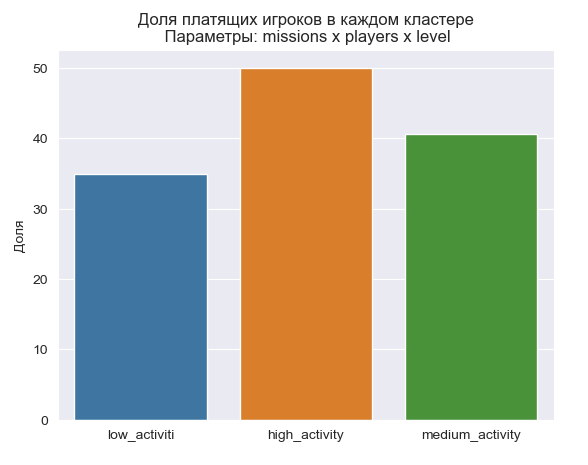




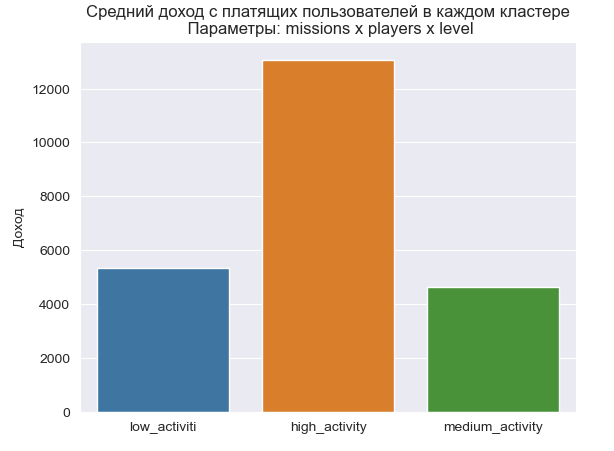
В то время как между кластерами низкой и средней активности разница достигаемого уровня достигает лишь 1 уровень, разница между средней и высокой активности составляет 3 уровня.

**Гипотеза:** *большинство создаваемых гильдий не активны*.

**Ответ:** В данной категории большинство гильдий не активны или имеют низкую активность.



|  |  |
| --- | --- |
| **Кластер** | **share\_payers** |
| **low\_activity** | 34.99 |
| **medium\_activity** | 40.62 |
| **high\_activity** | 50.00 |



|  |  |
| --- | --- |
| **Кластер** | **ARPPU** |
| **low\_activity** | 5321.86 |
| **medium\_activity** | 4633.24 |
| **high\_activity** | 13064.90 |

**Гипотеза***: игроки, покупающие премиум товары, создают более активные гильдии*, поскольку, инвестируя в игру, пользователь больше ценит вложенные усилия.

**Ответ:** В данном случае, рассматривая все гильдии, имеющие не пустые рассмотренные параметры, платящие игроки создают самые активные гильдии. Но так же стоит обратить внимание, что гильдии с низкой активностью приносят больший доход, нежели со средней активностью.

**Итог:**

Исследуя все данные, разбив их на несколько выборок, разбив каждую на кластеры по различным параметрам, можно сделать следующие выводы:

* В первой выборке, гильдии, активно участвующие в войнах гильдий, обычно уделяют меньше внимания другим игровым активностям. Самые активные гильдии в данной категории не имеют платящих игроков, но при этом гильдии со средним рейтингом и низким/средним количеством выполненных миссий имеют наибольший средний доход за платящего пользователя
* Во второй и третьей выборках, большинство гильдий не активны или имеют низкую активность. Однако платящие игроки создают самые активные гильдии в обеих категориях. Также стоит отметить, что гильдии с низкой активностью приносят больший доход, чем гильдии со средней активностью
* Платящие игроки играют важную роль в создании и поддержании активности гильдий, независимо от их уровня активности. Кроме того, гильдии с разными характеристиками могут иметь разный уровень доходности, что может быть важным для оптимизации игрового процесса

В целом, большинство создаваемых гильдий не активны.

В большинстве случаев, игроки, покупающие премиум товары, создают более активные гильдии.

если первое предположение верно, стоит ли существенно повысить стоимость создания гильдии, в том числе сделать это премиум товаром (за реальные деньги).

Возможно, стоит поднять стоимость гильдии, чтобы игроки видели больше ценности в ней, но точно не стоит её делать премиум товаром. Гильдии это одна из основополагающих аспектов в ММОРПГ и такое решение существенно снизит количество создаваемых гильдий, желание игроков их создавать и вступать в них, а соответственно, снизится количество игроков и все зависящие метрики.