***Шевченко Юлія Володимирівна***

**Завдання 1**

На мою думку, ці параметри будуть відображати поведінку користувачів у грі:

Payments – Кількість платежів. Подивимось наскільки гравець залучений;

Money – Загальна сума платежів. Показник монетизації;

LifeTime – Довжина життя користувача від моменту реєстрації до останньої активності, показує лояльність користувача на скільки він задоволений продуктом та готовий платити за додаткові функції підписки;

LastLong – Дані з моменту останнього входу. Оцінка активності користувачів чи залишаються вони залученими.

LastPayment – Дані з моменту останнього платежу. Активність платежів.

Sessions –Загальна кількість входів. Інтенсивність використання гри.

**2). Проведіть сегментацію. Визначте оптимальну, на вашу думку, кількість сегментів.**

Пропоную провести сегментацію кластеризації методом **K-Means, оскільки це поширений підхід для задач поділу ну групи. Я обрала розділити дані на 3 сегменти: Money**, **LifeTime** та **LastLong**. Тому що це активність, тривалість використання гри та монетизація.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cluster | Num\_Users | Total\_Money | Avg\_LifeTime | Avg\_LastLog | User\_Share | Money\_Share |
| 0 | 126 | 12402,94 | 640 | 7 | 12,6 | 9,21 |
| 1 | 632 | 56723,52 | 209 | 4 | 63,2 | 42,13 |
| 2 | 12 | 50599,68 | 295 | 4 | 1,2 | 37,58 |
| 3 | 230 | 14927,97 | 241 | 21 | 23 | 11,09 |

* **Кластер 0:** Помірна активність, довготривалі користувачі з низькими витратами.
* **Кластер 1:** Активні, новачки з середнім рівнем витрат.
* **Кластер 2:** Маленька група "VIP-користувачів" із найвищими витратами.
* **Кластер 3:** Застарілі або менш активні користувачі.

1. **Користувачі з помірною активністю:** 126 користувачів (12,6%), платежів 12 402,94 грн (9,2% від загальної суми). Висока тривалість перебування у грі (640 днів) при низькому рівні витрат. **Стратегія:** Стимулювати витрати за рахунок внутрішньоігрових бонусів чи персональних акцій.
2. **Активні користувачі:** 632 користувачів (63,2%), платежів 56 723,94 грн (42,13% від загальної суми). Найбільша група користувачів. Середня тривалість перебування грі (209 днів) при низькому рівні витрат. Дуже активні (остання активність – 3,6 днів тому).**Стратегія:** Тут можу запропонувати знижки щоб збільшити платежі гравців.
3. **VIP-користувачі:** 12 користувачів (1,2%), платежів 50 599,68 грн (37,58% від загальної суми). Найменша група користувачів,але які витрачають найбільше. Середня тривалість перебування грі (295 днів). Активні, із середнім часом останнього входу 3,8 днів. **Стратегія:** Залучення до ексклюзивного контенту або VIP-програми.
4. **Не активні користувачі:** 230 користувачів (23,0%), платежів 14 927,97 грн (11,1% від загальної суми). Середня тривалість перебування грі (241 день). Не активні, із середнім часом останнього входу 20,7 днів. **Стратегія:** Реактивація через бонуси або спеціальні пропозиції для повернення в гру.

|  |  |
| --- | --- |
| Statistic | Value |
| Average Age | 1959 |
| Median Age | 1959 |
| Most Common Age | 1970 |

Середнє значення віку у вибраних даних – 1959. Медіана також 1959. Але ту слід зауважити що середнє значення вони округлене в більшу сторону тому вони одинакові.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| User\_id | Year | Sex | Payments | Money | LifeTime | LastLog | LastPayment | Sessions |
| 589 | 1905 | male | 87 | 4236,2 | 833 | 1 | 7 | 1127 |
| 539 | 1905 | female | 5 | 6,97 | 820 | 1 | 1 | 10 |
| 345 | 1905 | male | 12 | 486,88 | 695 | 7 | 219 | 22 |
| 265 | 1910 | female | 3 | 2,97 | 458 | 1 | 383 | 1168 |
| 117 | 1905 | male | 2 | 5,98 | 436 | 15 | 360 | 223 |
| 18 | 1905 | male | 11 | 229,89 | 253 | 1 | 168 | 284 |
| 859 | 1910 | male | 5 | 30,95 | 246 | 1 | 133 | 135 |
| 795 | 1910 | female | 15 | 37,86 | 209 | 5 | 94 | 80 |
| 552 | 1915 | male | 1 | 2 | 68 | 1 | 40 | 159 |
| 163 | 1911 | female | 8 | 101,94 | 50 | 1 | 27 | 408 |

По значенням віку гравців які є аномальними тут я знаходила ці роки за методом квартелів. Цей метод підходить для виявлення аномалій на основі статистичних меж. Також слід поставити обмеження до 1920 року.