

Код python на [GitHub](#)

Введение

Я провожу исследование мобильной карточной игры.

Объектом исследования является поведение игроков во время игровых сессий.

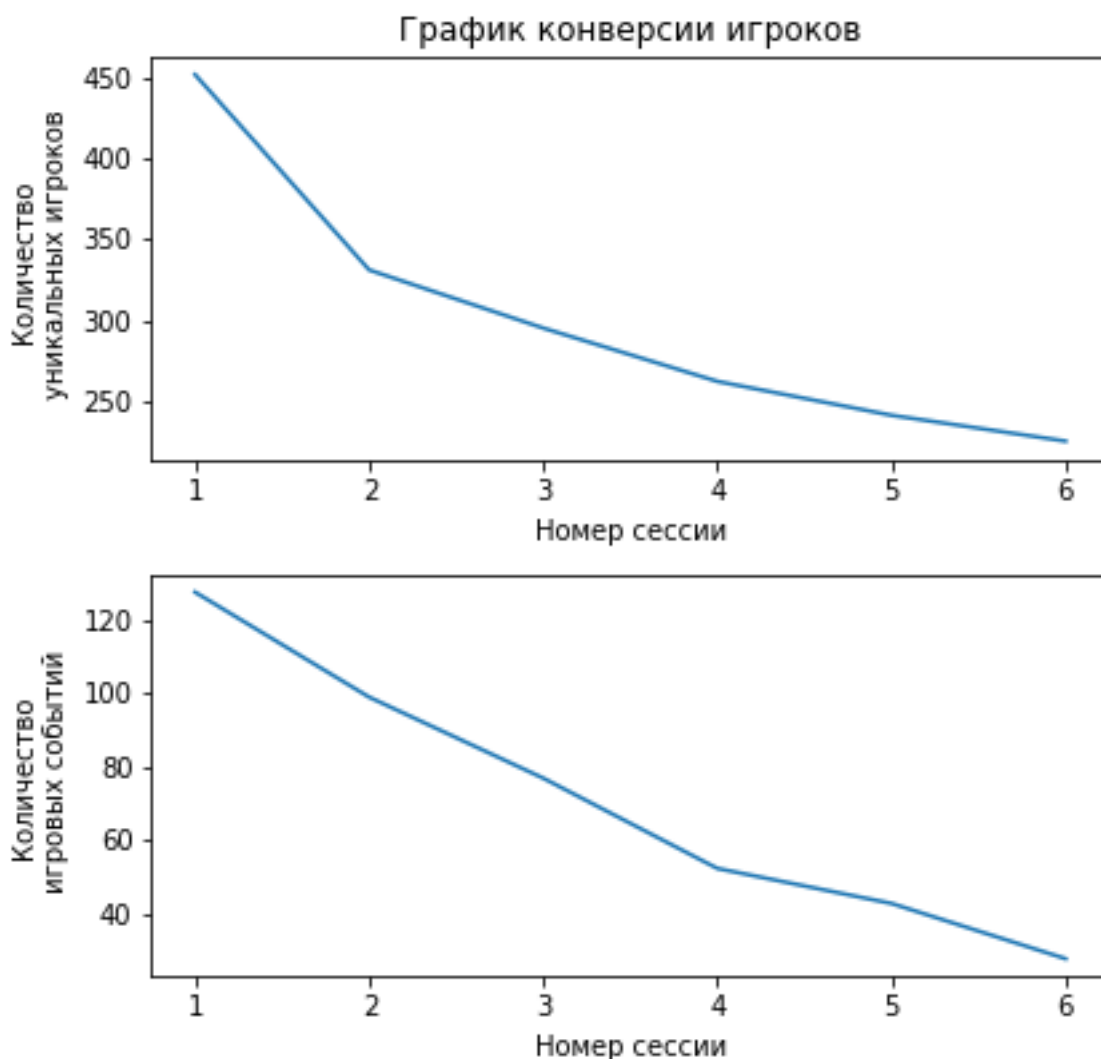
Задачей исследования является поиск аномалий в поведении игроков, объяснение этих аномалий и предложение путей решения.

Научно-исследовательской базой являются данные об игровых событиях.

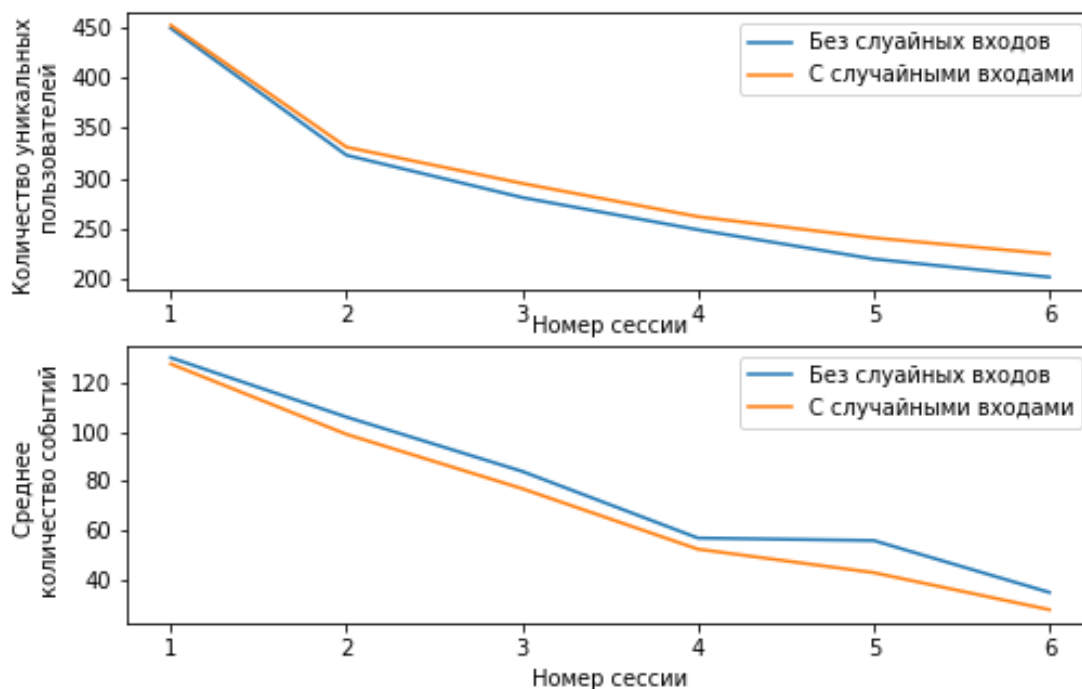
Перед проведением анализа необходимо было подготовить данные. Столбцы «**game_status**» и «**event_data**» содержали стандартизированные данные в формате « имя: значение », причем для каждого значения столбца «**event_name**» стандартная форма записи в столбце «**event_data**» отличалась. Необходимо было «раскрыть» данные по каждому событию, записать их в отдельный DF, чтобы позже объединить. Идея заключалась в том, чтобы после всех операций, сгруппировать события по столбцу «**user_id**», чтобы получить пользовательский путь в игре, с которым можно работать.

1.1 Первым пунктом для анализа является воронка обучения и вступления в игру (4 дня с подготовкой данных)

Я хочу посмотреть, сколько уникальных пользователей остается и сколько событий в среднем происходит с каждым последующим входом в игру для первых 6 сессий.

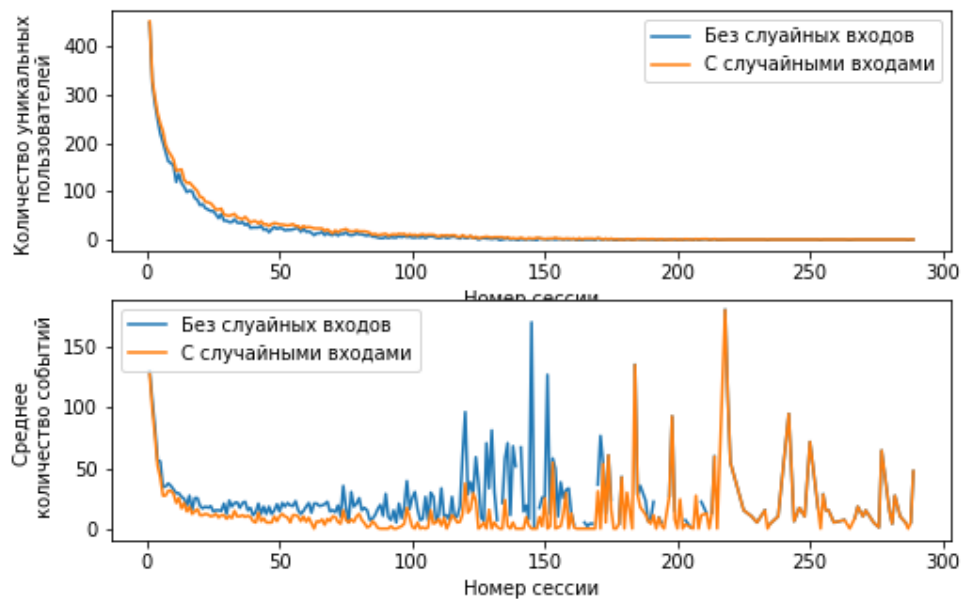


Как можно заметить на верхнем графике, линия графика изламывается больше всего на второй сессии. Было бы логично предположить, что среднее количество игровых событий пропорционально количеству уникальных игроков. Однако на нижнем графике такого излома нет. Возможно, данные искажены, из-за игроков, которые заходили в нее, но не играли (назовем это «случайным входом»). Попробуем убрать таких людей из общей выборки (количество игровых событий у них будет равно 1).



Действительно, это искажает данные, но можно констатировать, что большая часть игроков не запускает игру второй раз. Что касается игровых событий, к четвертому входу среднее их количество снижается практически линейно, но на пятый вход сохраняется неизменным, и падает к шестому (нижний график, синяя линия). Кроме того, можно заметить, что количество людей, которые осуществили «случайных вход», увеличивается с каждой новой сессией.

«Случайный вход» можно обосновать тем, что человек заходит, забирает ежедневный бонус, но в игру не играет. На более полном графике можно увидеть, насколько это явление искажает общую картину.



После 150+ сессии, большая часть людей входят в игру и совершают только одно действие.

После анализа графиков возникли следующие вопросы:

- Какие факторы из тех, что можно измерить численно, наиболее важны для того, чтобы игрок зашел в игру еще раз;
- Какое количество игровых событий нормальное (сколько игровых событий должно случиться, чтобы игрок с большей вероятностью зашел в игру еще раз?) и какие это события;

1.1.1 Какие факторы из тех, что можно измерить численно, наиболее важны для того, чтобы игрок зашел в игру еще раз

Идея заключается в том, чтобы найти с помощью деревьев решений (коэффициента Джинни) наиболее важные факторы (события, какие-то численные атрибуты типа среднего количества заработанных монет), которые наиболее важны для игрока.

Посмотрим на важные факторы для игроков, которые зашли в игру второй раз:

Сессия 1 \rightarrow 2

	feature	importance
0	freeRounds_game_status	0.217631
1	from_getCoins_Harvest	0.188469
2	name_learning_CRUISE_GET_CARD_FROM_DECK_MORE	0.160956
3	from_getCoins_LevelStart	0.158712
4	event_name_levelEnd	0.140541
5	prediction_levelEnd_VictoryWith4OrMoreCards	0.133690

Из данной таблички можно понять, что, например:

- игроку важно, какое количество бустеров с freeRounds у него в среднем за первую сессию;
- игрок, который прошёл хотя бы одну грядку с большей вероятностью вернется в игру;
- важно, какое количество подсказок о том, чтобы игрок взял дополнительные карты из колоды (хотя тут может играть роль и внимательность игрока);
- важно, сколько уровней игрок начал и закончил;
- важно, чтобы игрок, побеждал с 4 и более картами.

Сами по себе эти значимость факторов имеет мало практической пользы, однако, это отличные поинты для дальнейшего анализа.

Кроме того можно посмотреть важные факторы для последующих входов:

- Сессия 2 → 3

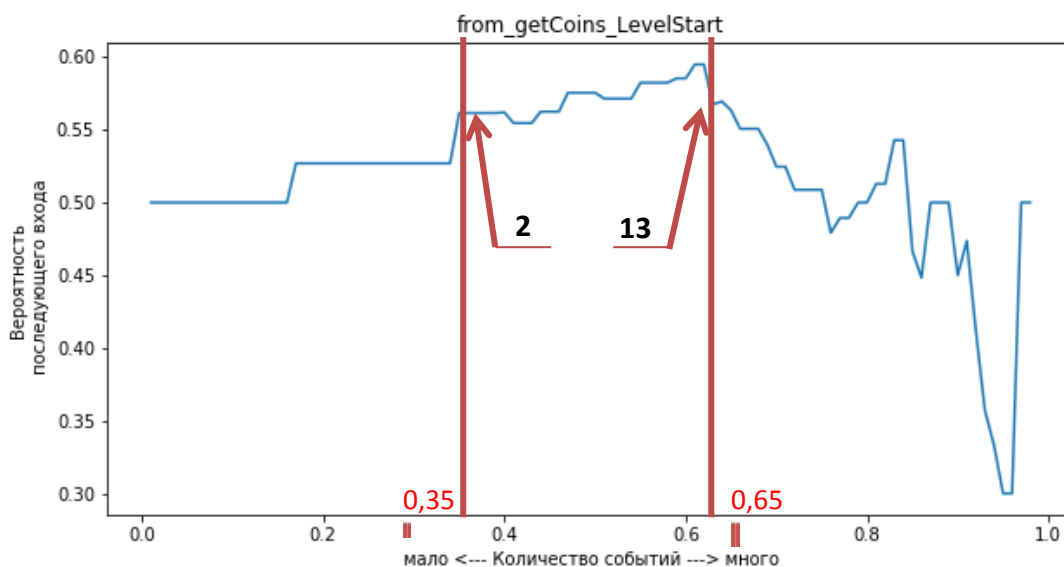
	feature	importance
0	from_getCoins_LevelBonus	0.269653
1	coins_game_status	0.245589
2	level_game_status	0.200935
3	freeUndo_game_status	0.178825
4	from_getCoins_AddCards	0.104997

- Сессия 3 → 4

	feature	importance
0	coins_game_status	0.243872
1	freeAddCards_game_status	0.195015
2	freeWildCard_game_status	0.157589
3	from_getCoins_UndoCardFromDeck	0.149641
4	prediction_levelEnd_DefeatWith1Or2Cards	0.135990
5	freeRounds_game_status	0.117892

1.1.2 Какое количество игровых событий нормальное (сколько игровых событий должно случиться, чтобы игрок с большей вероятностью зашел в игру еще раз?) для важных событий

Для лучшего понимания зависимости количества каждого события и вероятности второго входа в игру необходимо обратиться к следующим графикам:



Возьмем график, который проще понять и сформулируем гипотезу на основе этого графика. Вероятность повторного входа максимальная, если игрок сыграл от 2 до 13 игр в свою первую сессию.

Сформулируем гипотезы:

- Н0: Конверсия игроков, которые завершили от 2 до 13 уровней выше, чем у остальных игроков;
- Н1: Конверсия игроков, которые завершили от 2 до 13 уровней, такая же, как и у остальных игроков.

Тогда:

	Количество наблюдений	Количество повторных сессий	Конверсия
Группа Н0	160	126	0.7875
Группа Н1	292	205	0.702055

Для проверки статистической используется тест Фишера:

`z: 1.9620119622844294 p_value: 0.049761091568244174`

Для подтверждения нулевой гипотезы статистическая значимость «p_value» должна быть больше 0,05 для доверительной вероятности в 95%.

Нулевая гипотеза отвергается, принимается альтернативная гипотеза:

Конверсия игроков, которые завершили от 2 до 13 уровней, такая же, как и у остальных игроков.

Гипотеза не подтвердилась, возможно, потому, что я брал «количество уровней» как независимое событие. Это, конечно, неправильно. Нужно формулировать условие еще и на основании остальных графиков.

Стоит отметить, что для более точного определения вероятностей необходимо больше данных.

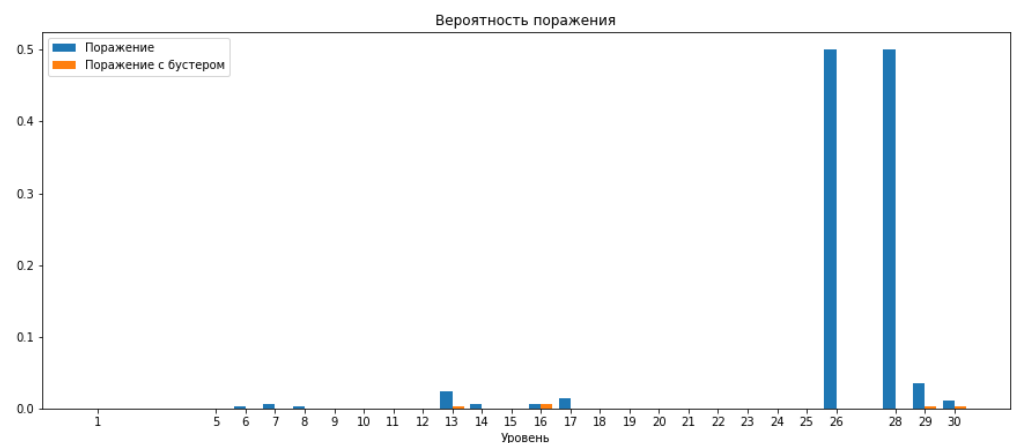
*Остальные графики можно посмотреть в приложении.

1.2 Вторым пунктом для анализа являются первые 30 уровней игры. (4 часа)

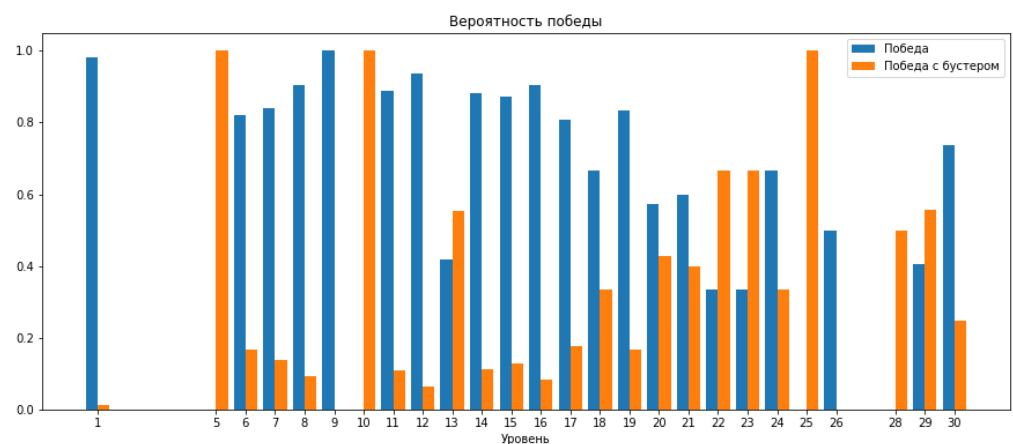
Необходимо определить, какие из них хорошие, какие нет и почему.

Для начала необходимо разобраться, какие критерия качества уровня, т.е. что такое «хороший» уровень.

Если говорить, что уровень плохой, потому что игроки проигрывают на нем, то выделяются уровни 26 и 28:

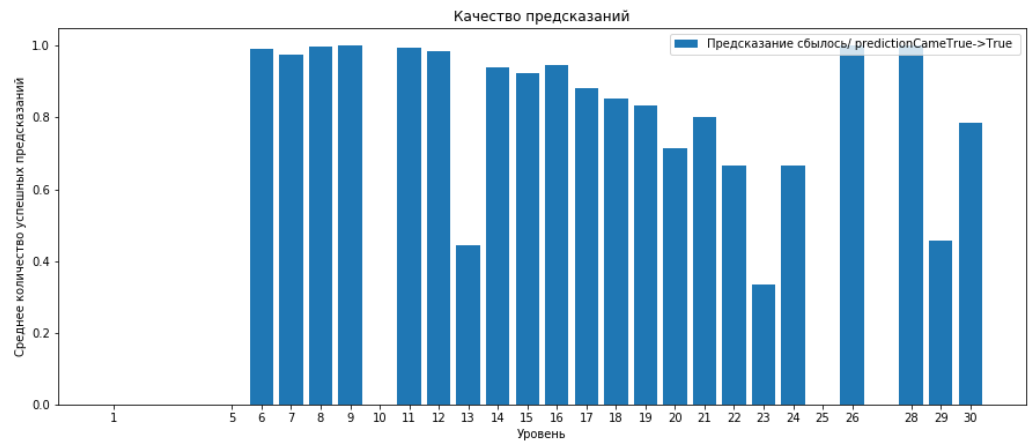


Если говорить, что уровень хороший, потому что игроки побеждают или побеждают с бустером:



На графике можно увидеть, что большая часть игроков на уровнях 13, 22, 23 и 29 скорее побеждают с использованием бустера. На уровнях 5, 10, 25 и 28 только с использованием бустера.

Если говорить о качестве предсказаний:



Для 1, 5, 10 и 25 уровней предсказания никогда не сбываются. В 13, 23 и 29 – сбываются редко.

1.3 Третьим пунктом анализа является эффективность рекламной компании, если затраты на одного пользователя равны 1.8 USD (1 час)

	Количество игроков	Стоимость одного игрока USD	Сумма рекламной компании USD	Количество игроков, которые совершили покупку	Сумма покупок USD	Эффективность рекламной компании
ad_traffic	159	1.8	286.2	7	49	0.171209

Как видно из данной таблички – с одного доллара затрат на рекламу возвращается 0,17 доллара прибыли.

Задание 2

В целом игра мне понравилась. Задумка и реализация очень хорошая. Графика на уровне других похожих проектов.

Мне хотелось бы увидеть обучение в виде небольших заданий, например: «сыграйте 2 игры» или «заработайте 10 звезд». За выполнение этих мини заданий, можно выдавать золото и бустеры.

Кроме того, есть проблема, что в игру перестают играть после n-ного количества сессий. Возможно, можно было бы сделать ежедневные задания на манер ЛБЗ из World of Tanks.

Я не заметил в игре чего-то такого, почему в нее стоит заходить каждый день и играть на протяжении долгого количества времени. Возможно, стоит оформить грядку как в «веселой ферме», чтобы игрок выбирал, что ему построить. Ему также придется набирать звезды, чтобы что-то построить. Но тогда можно будет продавать украшения для фермы или что-то подобное.

Возможно, стоит добавить таблицу лидеров.

Я не увидел кнопки «Магазин» когда играл.

Возможно, стоит добавить в игру аниме девушек, это сейчас модно. Да и можно продавать костюмы для них.

Приложение

