**Анализ мер по удержанию пользователей в мобильной игре “My café: Recipes and Stories” на рынке США**

**Аннотация**

В данной дипломной работе для анализа используется открытый датасет в виде тестовых данных, выгруженный из аналитической системы devtodev (<https://www.devtodev.com>). Датасет представляет собой данные о когорте пользователей, которые установили себе игру с 3 по 5 мая 2023 года (клиенты из России, Турции, США и Китая), а именно информацию об их игровой деятельности до 13 мая 2023 года включительно. Выбранная для анализа игра – “My café: Recipies and Stories”.Жанр игры – бизнес-симулятор ресторана, где игроку предлагается обустраивать и развивать свое кафе, получая прибыль от деятельности.

Модель монетизации игры – Free-to-Play. Это означает, что игру можно скачать и играть в нее бесплатно, однако для покупки эксклюзивного инвентаря и опций в игре, необходимо приобретать специальную валюту в виде бриллиантов. Их можно получить путем выполнения заданий, либо приобрести в магазине за реальные деньги.

1. **Анализ имеющихся данных, выявление узких мест.**

Для анализа данных использовались следующие инструменты:

1. Jupiter Notebook (Python) – очистка данных, расчет основных метрик, группировка данных по разным критериям, поиск инсайтов и узких мест в данных.

Файл в формате ipynb прилагается к отчету.

1. Tableau – дашборд с интерактивной визуализацией данных.

Ссылка на дашборд:

[My cafe. Общий анализ рынка](https://public.tableau.com/views/Mycafe_/Dashboard1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)

Первым этапом является подготовка данных к анализу в python. Я загрузила файл в формате csv, привела формат данных к нужному виду, избавилась от данных в некорректном формате, а также добавила несколько новых столбцов с важными для анализа показателями.

Как результат, адаптированный для работы датасет содержит следующие столбцы:

* Client ID – Уникальный клиентский идентификатор
* Install date – Дата установки игры пользователем
* Last seen – Последнее посещение пользователем
* Country – Страна пользователя
* Payer – Статус пользователя (платит/не платит)
* Paying capacity – Категория плательщика
* Predicted Paying status – Вероятность того, заплатит ли пользователь в будущем
* Payments sum – Сумма платежей
* Payments count – Количество совершенных платежей
* First payment – Дата первого платежа
* Last payment – Дата последнего платежа
* Channel – Источник прихода пользователя
* User level – Уровень игры, на котором сейчас находится пользователь
* Platform – Системная платформа пользователя
* Device – Устройство с которого пользователь заходит в игру
* Current app version – Текущая версия игры
* First app version – Первоначальная версия игры при установке
* Agency - Агентство
* Max days – Максимальное количество дней подряд в системе
* Push available – Статус пуш-уведомлений

Для проведения более корректного анализа, количество игроков из всех стран взято в одинаковом количестве.

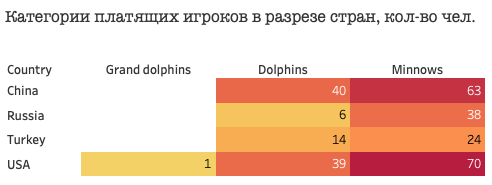
Прежде всего, были проанализированы клиенты по статусу платежности в разрезе стран. По каждой стране, платящими игроками являются не более 5% от общего количества игроков каждой страны. Наиболее высокие показатели в общей картине проглядываются у США и у Китая:



В аналитике игр существует разделение клиентов на категории по суммам, которую они готовы платить в игре:

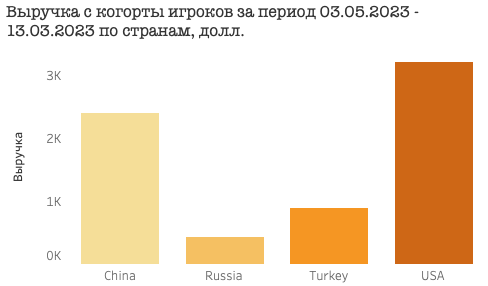
* Мальки (Minnows)
* Дельфины (Dolphins)
* Киты (Whales)

Иногда, для более детальной сегментации, аналитики добавляют промежуточные значения. В нашей игре есть категория больших дельфинов (grand dolphins):

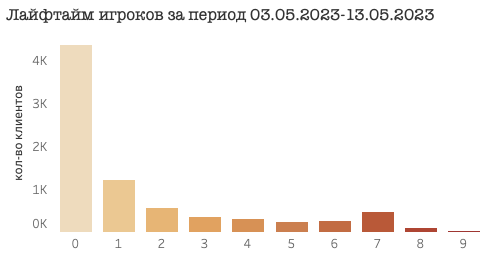


Итак, рассматривая таблицу в разрезе стран, мы можем увидеть, что в категорию больших дельфинов попал только один игрок из США. Как было отмечено и при анализе предыдущего графика, игроки из США и Китая готовы платить больше, чем из Турции и России.

Далее рассмотрим общие финансовые показатели в разрезе стран. Картина вполне сопоставима с анализом предыдущих двух графиков – самой прибыльной страной является США, далее по убыванию следуют Китай, Турция и Россия. Выручка от российских игроков более чем в 6 раз меньше в сравнении с выручкой от игроков Соединенных Штатов.

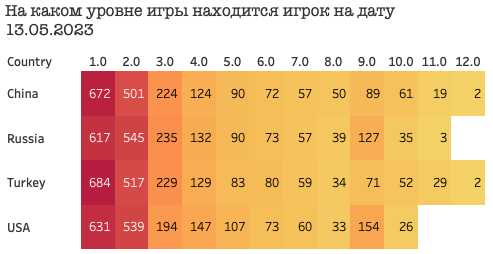


Теперь рассмотрим подробнее насколько нашему приложению удается удерживать клиентов в системе, то есть продолжительность пребывания игроков в системе (в днях):



Мы видим, что огромный процент пользователей ушел из системы сразу в день установки приложения. В первый день также прослеживается заметный отток (примерно в 2 раза).

Еще одна интересная группировка пользователей – это понимание на каком уровне игры пользователи находятся на день анализа датасета:

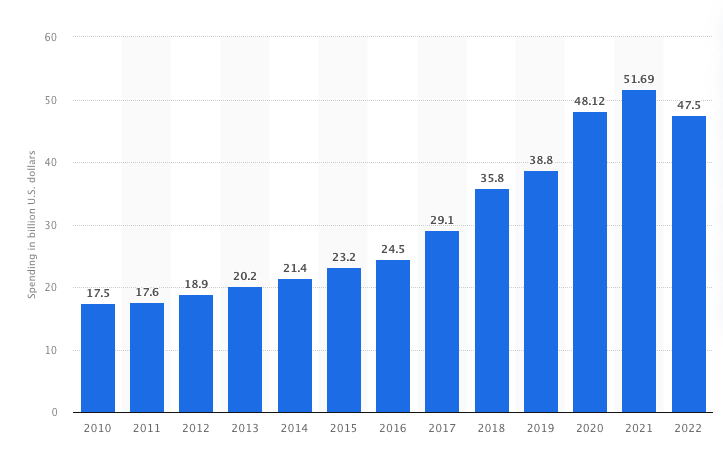


Мы видим, что в большинстве своем пользователи забрасывают игру на 1 и 2 уровнях, что коррелирует с оттоком пользователей уже в первый день после установки приложения. Проблема особенно выделяется на рынке США – тут мы видим самый низкий показатель удержания пользователей после 2го уровня игры.

Итак, одной из замеченных закономерностей является то, что многие клиенты остаются на уровне 2 (в начале игры), и на этом уровне происходит достаточно большой отток игроков. Сделав сегментацию по странам, я выявила что проблема особенно выделяется среди игроков из США. При этом, процент платящих клиентов в этой стране самый высокий, как и выручка с этих клиентов, что показывает хорошие перспективы при увеличении уровня удержания клиентов. С этой точки зрения было принято решение рассмотреть рынок США на более детальном уровне.

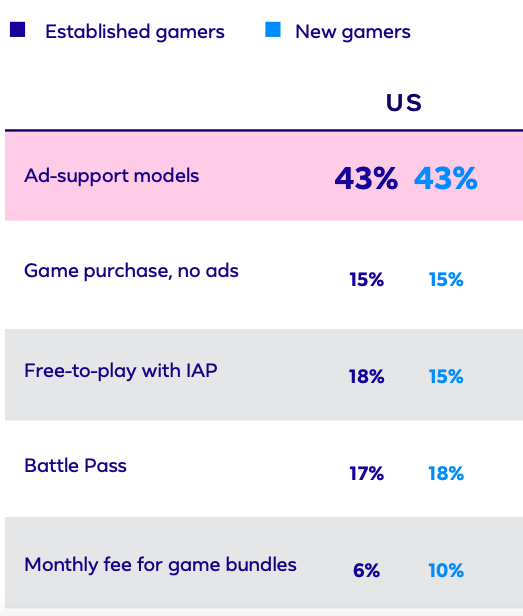
1. **Исследование рынка мобильных игр**
   1. **Тренды на рынке мобильных игр США. Основные конкуренты.**

Рассмотрим график выручки от видеоигр по рынку США за последние 12 лет:

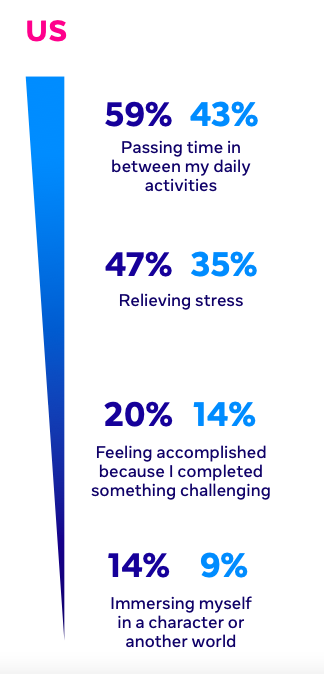
****

Можно увидеть, что в период с 2010 по 2019 год наблюдался постепенный рост выручки, однако в 2020 году мы видим значительный скачок вверх. Он обосновывается активным периодом пандемии. Активные вложения потребителей в игорную индустрии растянулись на 2020 и 2021 год, в 2022 году же наблюдалось небольшое снижение ниже уровня 2020 года вследствие затихания пандемии. В связи с активным развитием технологий виртуальной реальности и искусственного интеллекта, можно ожидать растущий спрос и, соответственно, увеличение выручки в ближайшие несколько лет.

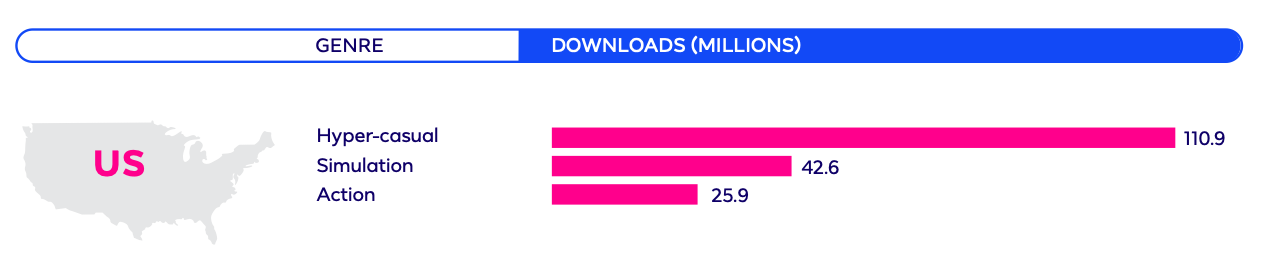
Что касается предпочитаемых типов монетизации для игроков, самым популярным типом остается рекламная модель, которая не накладывает на игрока обязательств осуществлять платежи. На втором месте – free-to-play игры (условно бесплатная игра со встроенными покупками), к которой как раз относится игра My Café:

****

Если говорить об основных мотиваторах, побуждающих игроков к игре на рынке США, то прежде всего это возможность скоротать время в перерывах между ежедневными активностями. Затем следует игра – как средство избавления от стресса:

****

Теперь рассмотрим наиболее популярные жанры среди игроков Соединенных Штатов. Топ-3 составляют гиперказуальные игры, симуляторы и экшн. Количество скачиваний игр-симуляторов в 2022 году составило 42.6 млн:

****

На рынке присутствует немало игр-симуляторов в области общепита. Основные из них:

# Restaurant Empire

# Burger Shop

# Baking Life

### Cake Mania: Main Street Lite

* 1. **Особенности игр-симуляторов**

Игры-симуляторы имитируют какую-либо деятельность реального мира. Игроки управляют действиями персонажей в игровом мире, выбирают, как реагировать на происходящие события, или управляют каким-либо устройством, например транспортным средством. Это отличает игры-симуляторы от традиционных аркад или головоломок, которые обычно имеют заданную последовательность задач и решений. Эти игры предлагают игрокам почувствовать контроль над окружающей средой, полностью погружая их в виртуальный мир.

Жанр стал особенно популярным в 1980-х годах. Аркады вводили симуляторы, чтобы предоставить игрокам более полный опыт. В то время как аркадные симуляторы предназначались в основном только для развлечения, игры для ПК были больше ориентированы на реализм и моделирование сценариев реального мира. Они использовались в качестве образовательных платформ, чтобы пользователи лучше понимали, как принимаются решения в реальном мире.

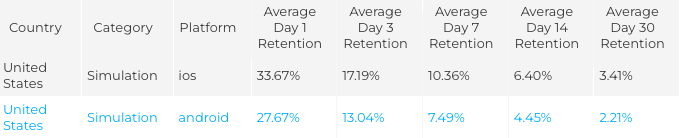
Со временем жанр развился и стал включать в себя больше фантастических и нереалистичных элементов. Например, The Sims — это игра-симулятор, которая позволяет игрокам управлять жизнями цифровых персонажей в смоделированном мире. RollerCoaster Tycoon — еще один пример, в котором игроки могут построить тематический парк и управлять им.

Жанр постоянно развивается. Постоянно внедряются новые технологии и игровые механики. Виртуальная реальность и искусственный интеллект играют важную роль в том, как симуляция будет использоваться в будущем. Постоянное совершенствование виртуальной реальности важно для симуляторов, поскольку дает игрокам возможность полностью погрузиться в игровой мир.

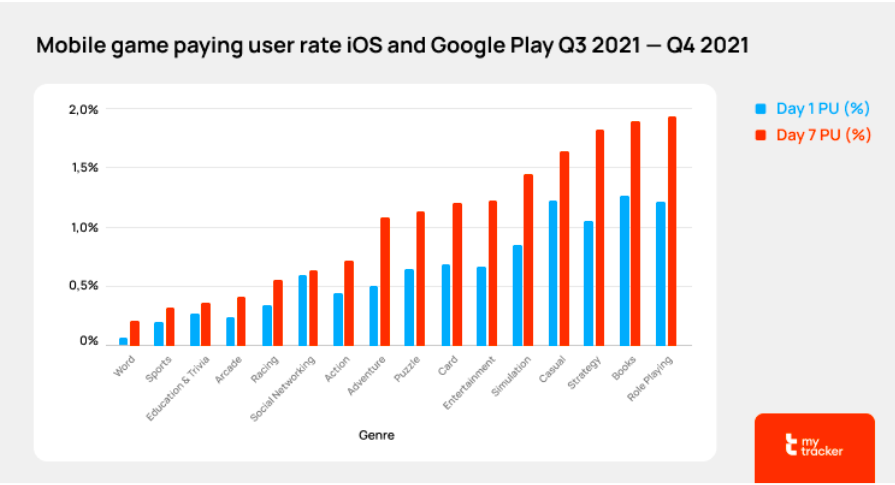
Игры-симуляторы также становятся все более популярными на мобильных устройствах. Мобильные симуляторы позволяют игрокам испытать множество различных симуляций на своих телефонах или планшетах. Использование мобильного устройства также позволяет выполнять некоторые симуляции в режиме реального времени, поскольку устройство всегда находится рядом.

**2.3 Бенчмарки по рынку**

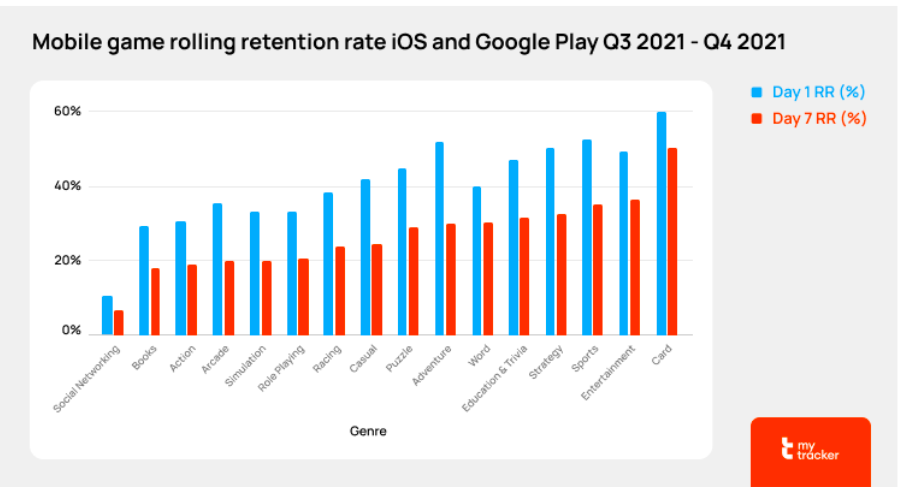
Для дальнейшего анализа нам необходимо выявление среднерыночных метрик удержания в сфере видеоигр. Рассмотрим метрики Retention игр-симуляторов в США, посчитанных по разбивке по платформам. Что интересно, Retention у игроков, использующих платформу Apple, выше в любой из дней. Средний Retention Day 1 составляет около 30%, Retention Day 7 – около 9%:



Теперь рассмотрим статистику платящих игроков на 1 и на 7 день после установки игры. В первый день более 0,8% всех любителей симуляторов готовы платить за игру, на 7 же день платежными становятся почти 1,5% игроков. В сравнении с другими жанрами, симуляторы по количеству платящих игроков занимают 5 место, уступая казуальным играм, стратегиям, книгам и ролевым играм.

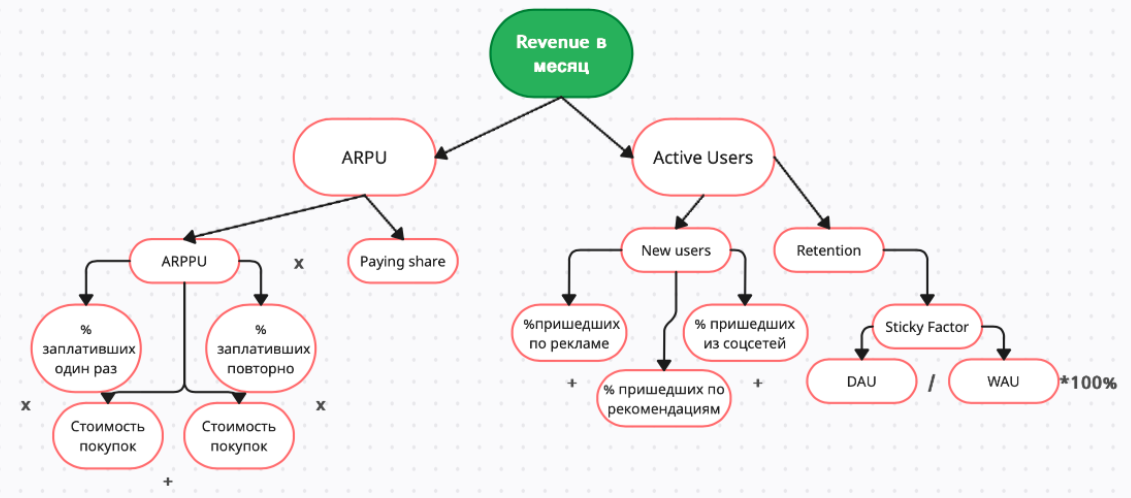


Теперь что касается метрики Rolling Retention Day 1 и Day 7 – у симуляторов оба этих показателя имеют не слишком высокие значения в сравнении с другими жанрами, здесь симуляторы занимают 5-ю позицию с конца.



1. **Работа с гипотезами**
   1. **Иерархия метрик**

Для того, чтобы определить основные зависимости между метриками, проанализируем зависимости и построим иерархию метрик:

****

Итак, во главе нашей структуры находится Revenue (Выручка), являющаяся основополагающей метрикой, фиксирующей финансовый результат функционирования нашей игры. Далее разделение происходит на 2 крупные метрики:

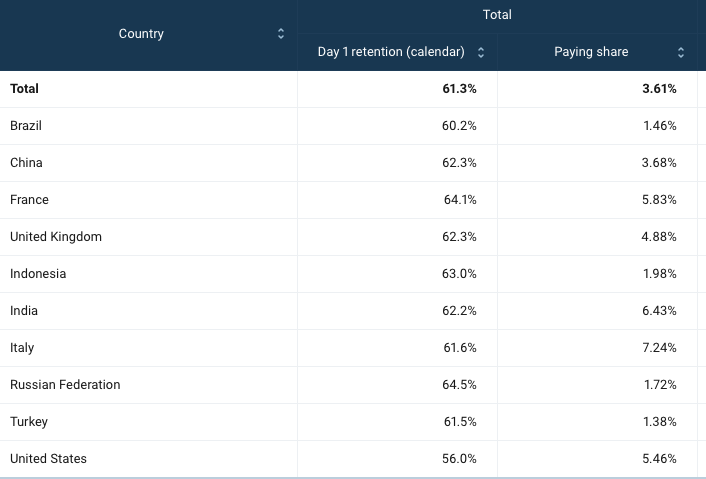
ARPU – average revenue per user, то есть средний доход, который нам приносит один пользователь приложения за определенный период. ARPU можно разбить на ARPPU (средний доход, который приносит пользователь, совершивший хотя бы один платеж за период) и долю платящих клиентов (paying share, %).

Active users - число уникальных игроков за месяц, у кого была хотя бы одна сессия за исследуемый период. Таких пользователей можно поделить на новых пользователей и тех, которые перешли к нам с предыдущего периода. Поэтому этих пользователей можно «покрыть» метрикой Retention, обозначающей удержание игроков в игре. Оно, в свою очередь, зависит, в первую очередь, от уровня вовлеченности пользователя в игру, что на языке метрик называют Sticky Factor.

* 1. **Формулирование гипотезы на основе выявленной проблемы**

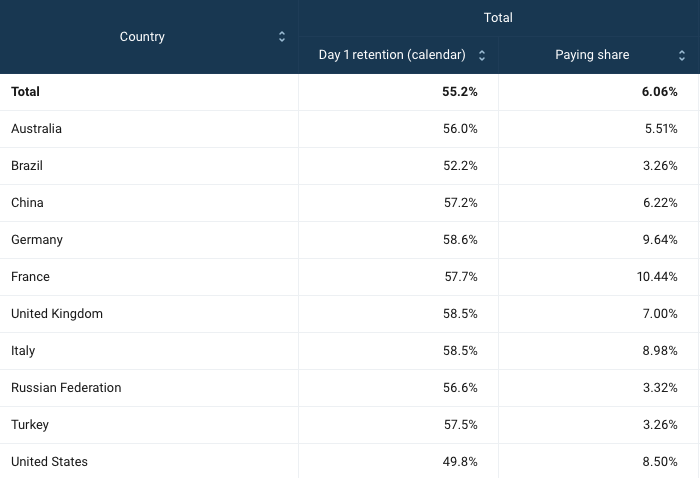
В первой части исследования мы выявили, что наиболее проблемным рынком с точки зрения оттока клиентов является рынок США, который, однако, является одним из самых прибыльных. Отчеты в системе dev-to-dev позволяют нам выявить данные по игре по метрике Retention и сравнить его с бенчмарками на рынке США:

1. Расчет метрики Retention 1 Day и Paying Share для игроков-пользователей Android:



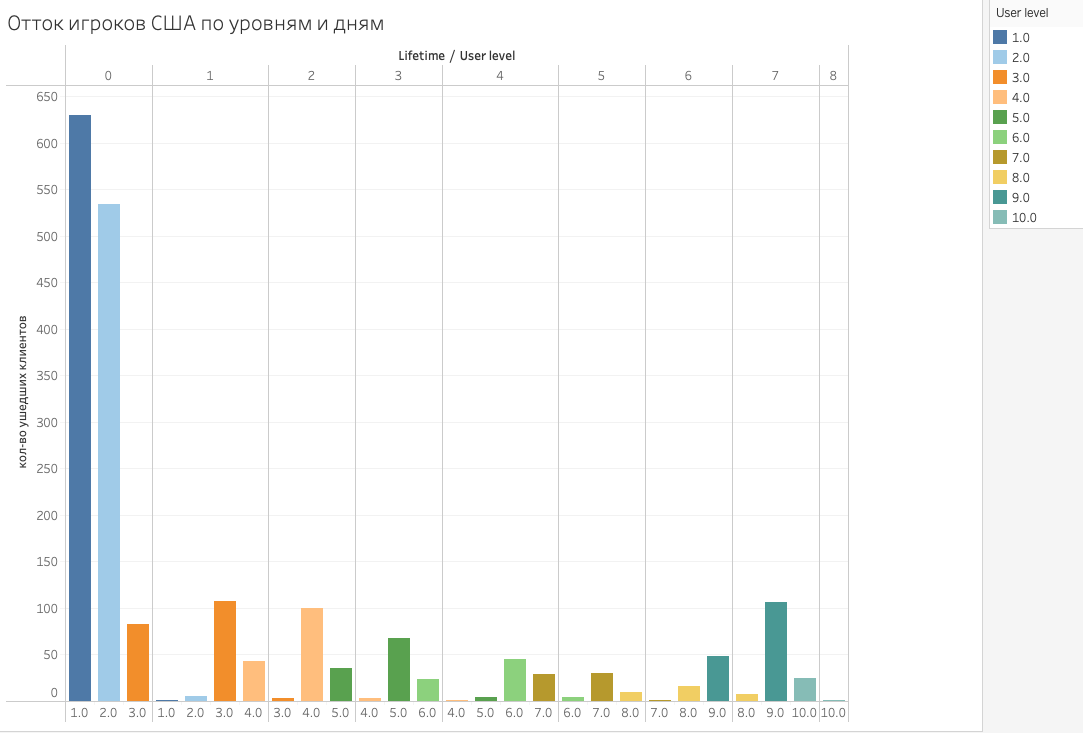
При сравнении с бенчмарками США, показатель удержания 1-го дня нас приятно удивляет, так как в 2 раза превышает средний показатель по рынку по играм-симуляторам. Однако, что бросается в глаза, процент удержанных в первый день пользователей в среднем на 5-6% уступает Retention в других странах. При этом, как можно увидеть, процент платящих игроков (Paying Share) в США является одним из самых высоких.

1. Расчет метрики Retention 1 Day и Paying Share для игроков-пользователей Apple:



Подобную ситуацию мы видит также и у пользователей платформы Apple в США. Доля платящих тут составляет практически 10% от общего количества игроков, однако процент удержания в первый день является самым низким.

Теперь рассмотрим график оттока игроков по дням в разрезе уровней:



Большинство клиентов в нулевой день ушли еще на первом (обучающем) уровне, возможно поняв, что эта игра им неинтересна. Однако, наиболее сильный разрыв наблюдается между вторым и третьим уровнем, причем основная масса покидает игру на втором уровне еще в день установки приложения. Можно сделать вывод, что вовлеченность игры падает именно на нем. Разобравшись в причинах и найдя способ увеличения вовлеченности на 2 уровне, мы можем положительно повлиять на метрику Retention 1-го дня.

Продуктовая цель: Повышение показателя Retention Day 1 по рынку США на 8% до конца июля 2023 года.

Выявленная проблема: Многие игроки бросают игру на уровне 2, что уменьшает лайфтайм и удержание клиентов в игре.

Возможная причина: однообразие 1го и 2го уровней. Если первый уровень игры мы можем рассматривать как обучающий, то во втором уровне игрокам начинают докучать однообразие игры и они уходят. При этом, во втором уровне игры клиентом еще не доступны большинство предметов в магазине, при том, что деньги (монеты) имеются у них на счету, игрок не имеет возможности персонализировать свое кафе, что может послужить причиной оттока. Простыми словами, игрок не чувствует свободы действий, а монотонность из уровня в уровень ему наскучивает.

Гипотеза: если мы введем по несколько доступных предметов в каждую подкатегорию в магазин на 2 уровне и сообщим игроку о возможности персонализации кафе на этом уровне, то метрика Retention 1-Day повысится на **8%.**

Метрики, требующиеся для оценки гипотезы: Retention 1st Day

Для проверки гипотезы предлагается провести A/B тестирование на выборке из свежей аудитории США, которая еще не успела дойти до 2 уровня игры.

* 1. **Подготовка данных к A/B тестированию**

**Суть эксперимента:** эксперимент будем проводить на аудитории из США, пришедшей за последние сутки, и которая еще не успела подойти к уровню 2 игры. Планируется разделение игроков на две однородные группы – тестовую и контрольную, для одной из которых мы откроем возможность разнообразить интерьер кафе на втором уровне (добавим ряд открытых предметов для покупки + уведомление о возможности воспользоваться этими предметами). Для контрольной же группы функционал останется прежним.

**Длительность эксперимента:** как мы можем увидеть на графике оттока игроков по дням, игроки доходят до 2 уровня уже в первые два дня после скачивания приложения. Поэтому эксперимент может занять у нас всего около трех дней, но для нас важно отследить, повлияют ли наши изменения на более долгосрочные метрики, например Retention 7-Day. Поэтому предлагается проводить эксперимент в течение восьми дней.

**Определение уровня значимости и мощности теста.**

В качестве подготовки к тестированию, нам необходимо учесть возможность появления ошибок 1го и 2го рода.

Ошибка 1го рода – это вероятность отклонения нулевой гипотезы, при том, что на самом деле она верна. За это отвечает уровень значимости. Определим его самым популярным значением 5%.

Ошибка 2го рода – это вероятность принятия нулевой гипотезы, при том, что она на самом деле ошибочна. За это отвечает мощность теста. Возьмем по умолчанию значение мощности 80%.

Определим минимально возможное значение выборки для проведения теста. Для этого воспользуемся онлайн-калькулятором. Все необходимые показатели мы определили:

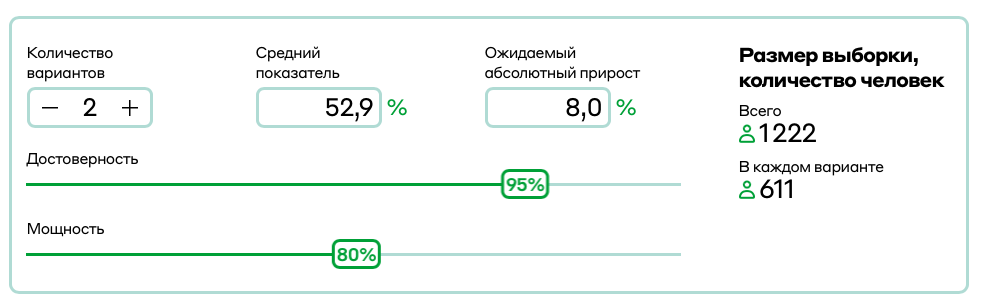
Количество вариантов (групп): 2 группы

Уровень значимости: 5%

Мощность теста: 80%

Средний показатель (текущий уровень Retention 1 Day): 52,9%

Ожидаемый абсолютный прирост (из гипотезы: 8%)



Таким образом, количество игроков для проведения эксперимента равно 1222. То есть в каждую группу мы можем включить по 611 человек минимум.

**Проверка исторических данных по метрике для выбора подхода тестирования**

Следующим этапом необходимо проанализировать нашу метрику на нормальность. Сделать мы это можем на основе исторических данных. Для этого мы можем взять выборку из средних значений Retention по входам в первый день после установки игры. Система dev-to-dev позволяет нам получить эту информацию благодаря своим интерактивным отчетам.

Соответствующие расчеты по метрике Retention 1 Day я проводила в Jupiter Notebook ( ноутбук Bootstrap прикладываю к дипломной работе). Была возможность использовать либо критерий Шапиро-Уилка, либо метод бутстрепа для проверки метрики на нормальность. Я остановилась на использовании бутстрепа. Суть метода заключается в том, что мы «тиражируем» исходную выборку очень большое число раз и вычисляем средние значения на основе отобранных значений. Процедура «тиражирования» может производиться 500-10000 раз. Далее, исходя из нашей цели, на базе полученных значений мы проверяем данные на нормальность.

В нашем случае мы произвели «тиражирование» 5000 раз. На основе полученных значений была построена гистограмма:

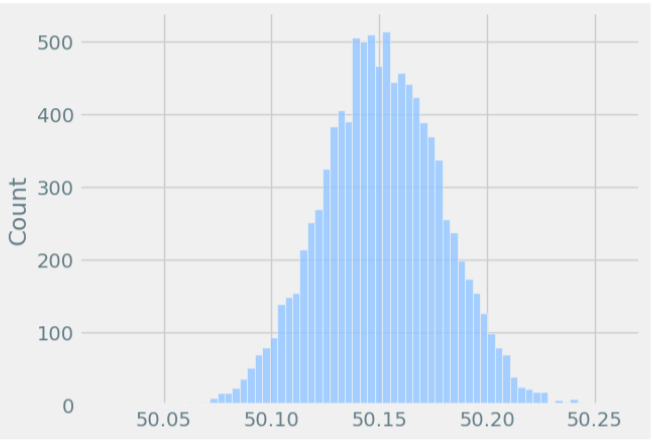
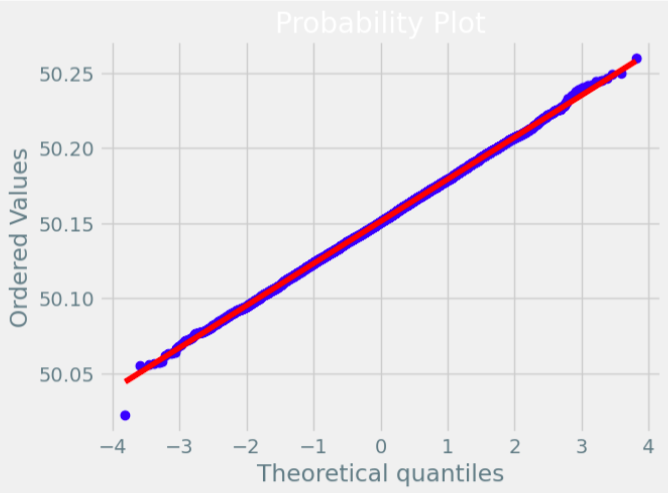


Диаграмма имеет форму достаточно ровного колокола, что характеризует нормальность распределения данных. Для более точной интерпретации обратимся к Q-Q графику. Чем ближе точки на графике к диагональной прямой, тем ближе распределение исследуемой переменной к нормальному закону:



Распределение наших данных практически совпадает с диагональной прямой, что еще раз характеризует их его нормальность, а все точки (за исключением одной) расположены на или очень близко к линии, что говорит о гомогенности дисперсии.

Исходя из проведенного анализа, мы можем сделать вывод, что для анализа результатов теста по метрике retention 1 day мы можем использовать ttest Стьюдента.

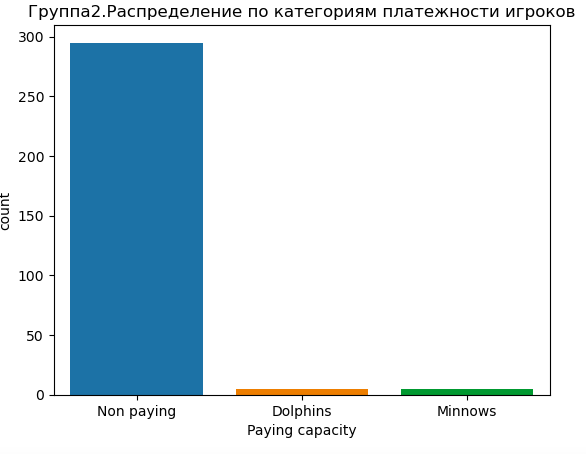
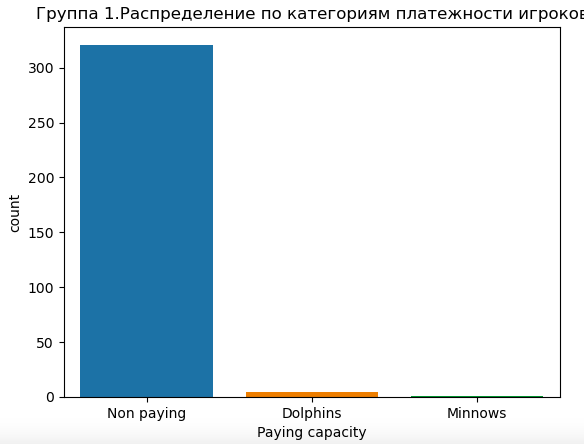
**Формирование выборки, сплитование на группы и проверка на однородность.**

Теперь нам необходимо сделать проверку выборки для теста на однородность. Мы будет осуществлять проверку также на основе исторических данных – точнее на основе датасета, который мы использовали для выявления «узких мест».

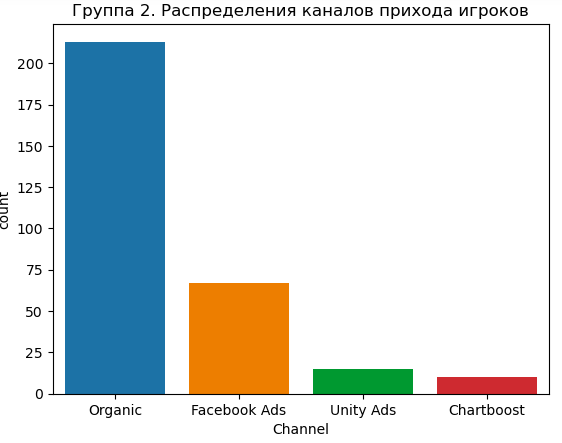
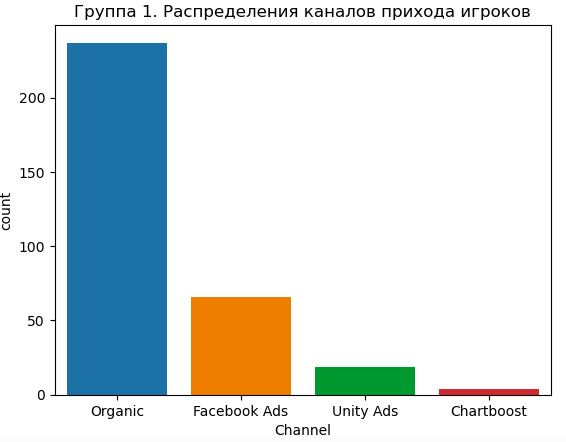
Первым делом я отфильтровала игроков по стране (США), затем по уровню, на котором сейчас находится игрок – были взяты только те игроки, которые на текущий момент не дошли до 2го уровня. Далее методом рандома выборка была поделена на 2 практически равные части.

Чтобы убедиться в однородности двух выборок, я графически отобразила результаты разбивки по нескольким категориям:

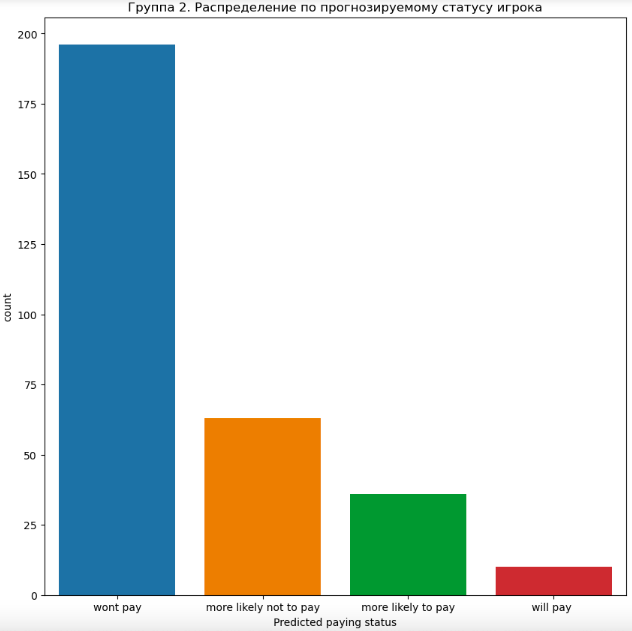
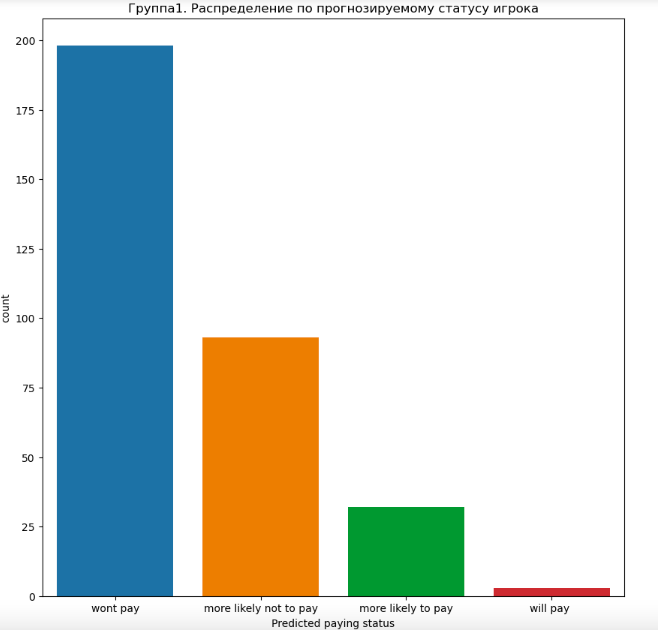
* Платежная группа игроков (неплатящие, мальки, дельфины)



* Источники прихода игроков



* Прогнозируемый статус клиента (будет/не будет платить в будущем).



Из сравнений графиков двух групп по трем показателям, мы можем сделать вывод что данные выглядят вполне однородными и можно продолжать проведение A/B тестирования. Для более тщательной проверки на однородность рекомендуется проведение A/A теста.

**Выводы**

* Видимое «узкое место» обнаружено на рынке США – наиболее сильный отток игроков после 2го уровня еще в нулевой день, по сравнению с другими странами;
* При этом доля платящих игроков и выручка от рынка США характеризуются наиболее высокими показателями;
* Рекомендуется провести А/В тест с тестируемой метрикой Retention 1 Day, введя изменения в сценарий 2го уровня игры; данные для анализа подготовлены и протестированы на нормальность.