

# Тестовое задание

## на позицию «Игровой Аналитик Данных (gamedev)»

### 1. Мобильные проекты, в которые вы играли

Перечислите мобильные проекты, в которые вы играли за последние полгода на телефоне (планшете).

По указанным проектам укажите, сколько денег они заработали за 2023 год.

### 2. Разбор игры

Выберите одну из игр из п.2 и проведите её разбор. Вот пример разбора, который нам нравится: [ссылка на пример](#).

План минимум для разбора:

- Опишите её основные механики, игровые циклы, модель монетизации.
- Порекомендуйте, как этой игре увеличить 1-day retention, 7-days retention, ARPU, как увеличить LTV игрока.
- Какие вопросы может в ежедневном формате задавать себе разработчик этой игры?
- Какие события для этого нужно передавать в аналитическую систему?
- Какие метрики выведете на основной дашборд?

### 3. Анализ снижения выручки

Представьте, что вы игровой аналитик, и на вашем проекте (игре) во вчерашний понедельник на 30% упала выручка к предыдущему дню. Вам необходимо выяснить причины снижения выручки, какие гипотезы вы бы стали проверять, каким образом и какие данные вам для этого потребуются.

### 4. SQL-запросы

Есть 3 таблицы:

Installs - таблица, содержащая информацию об установках:

```
install_time timestamp
client_type varchar(50)
country_code varchar(255)
user_id varchar(255)
```

Visits - таблица с визитами:

```
id int(10)
user_id varchar(255)
time timestamp
client_type varchar(50)
ip varchar(255)
```

Payments - таблица с информацией о платежах:

```
id int(10)
user_id varchar(255)
transaction_time timestamp
product_id varchar(255)
amount decimal(19,4)
```

Напишите SQL-запросы для расчета:

- ARPU, ARPPU, Paying share для пользователей, установивших приложение в последние 30 дней.
- RR1, RR7 (классический).
- Конверсию во 2-й платеж в течение 14 дней с момента установки для пользователей, установивших приложение в последние 30 дней.

## 5. Python

Для выполнения этого задания требуется сгенерировать DataFrame с синтетическими данными. DataFrame должен состоять из 1000 строк и 5 колонок. Каждую из колонок мы предлагаем создать и заполнить следующим образом:

- 1-я колонка – `user_id` – идентификатор пользователя. Длина `user_id` должна равняться 15-ти символам. Идентификатор состоит из случайной комбинации следующих символов: "1234567890abcdefghijklmnopqrstuvwxyz". Для каждой строки в DataFrame значение `user_id` формируется случайным образом. Добавьте проверку на уникальность сгенерированного `user_id`.
- 2-я колонка – `purchase_number` – номер покупки. Столбец необходимо заполнить случайными значениями в диапазоне от 1 до 10.
- 3-я колонка – `interval` – время, прошедшее с момента предыдущей покупки пользователя. Столбец необходимо заполнить случайными значениями из нормального распределения со средним 1440 и стандартным отклонением 200. Обратите внимание, что для первой покупки пользователей данное поле должно быть пустым (null), так как предыдущей покупки у пользователя не было.
- 4-я колонка – `amount` – стоимость покупки. Значения для этого столбца необходимо взять из экспоненциального распределения с параметром  $\lambda = 1$ .
- 5-я колонка – `retention` – день жизни покупателя, в который он совершил покупку. Необходимо сгенерировать значения 1, 2, 3, 4, 5 с вероятностями 0.35, 0.25, 0.2, 0.15 и 0.05 соответственно.

### Задача:

- Для всех строк исходного датасета, сгруппированных по номеру покупки, посчитайте среднее значение интервала от предыдущей покупки (3-я колонка) по группе. Результат необходимо добавить в новый столбец датафрейма.
- Напишите функцию, которая принимает на вход значение `user_id` и возвращает строку следующего вида: все буквы в той последовательности, в которой они встречаются в `user_id`, затем все цифры в `user_id` в той последовательности, в которой они встречаются в `user_id`. Добавьте результат применения этой функции к `user_id` в новый столбец.
- Вычислите моду, медиану, среднее, дисперсию и стандартное отклонение для столбцов `interval`, `amount` и `retention`. Постройте гистограммы распределения значений в столбцах.
- Постройте график, который наиболее полно, на ваш взгляд, описывает зависимость интервала между покупками и номера покупки.