1. **Создание ДФ**

Дата-фрейм будет состоять из 10000 строк и 5 колонок.  
Каждую из колонок мы предлагаем тебе создать и наполнить далее.

**1-я колонка** - user\_id. Длина user\_id должна равняться 15-ти символам и состоять из случайной комбинации следующих символов: "1234567890abcdefghijk". Комбинации каждого id в df формируются рандомно.

**2-я колонка** - номер заказа. Случайные значения лежат в диапазоне от 1 до 10.

**3-я колонка** - время доставки. Значения берутся из нормального распределения со средним 1440 и стандартным отклонением в 200.

**4-я колонка** - общая стоимость заказа. Значения берутся из экспоненциального распределения с параметром лямбда = 1, смещённого на +1.

**5-я колонка** - ретеншен - день жизни покупателя, в который он совершил заказ. Сгенерировать значения 1,2,3,4,5 с соответствующими вероятностями 0.35, 0.25, 0.2, 0.15, 0.05.

В случае дублей user\_id, оставьте только первую из дублирующихся записей.

1. **Дополнение ДФ**

* Добавьте новый столбец - для всех строк исходного дата-сета, сгруппированных по номеру заказа, присвойте среднее значение времени доставки по группе.
* Определите сегмент, к которому относится пользователь, на основании номера заказа, который у него указан. Определённый сегмент присвойте каждому пользователю в новом столбце. Сегмент состоит из диапазона заказов по 3. Итоговые лейблы сегментов в новом столбце должны выглядеть так: "1-3","4-6" и т.д.
* Берём группы по 2000 строк дата-фрейма, отбираемых по порядку. Переменную "Общая стоимость заказа" в группах последовательно перемножьте на 1,2,3,4,5. То есть в первой группе умножаем переменную на 1, во 2-й – на 2 и т.д. Полученные значения должны заменить прежние.

**ЗАДАНИЕ ПОСЛОЖНЕЕ - НЕОБЯЗАТЕЛЬНОЕ**

* Отдельной колонкой добавьте значения последовательности, начинающейся с 0,1, где каждый следующий элемент является суммой двух предыдущих, умноженных на 0.5

1. **Исследование ДФ**

* Посчитайте или покажите моду, медиану, среднее, дисперсию и стандартное отклонение для 4-й колонки - суммы стоимости заказа.
* Постройте график, который наиболее полно, на ваш взгляд, описывает время доставки заказа в зависимости от его номера.

**ЗАДАНИЕ ПОСЛОЖНЕЕ – НЕОБЯЗАТЕЛЬНОЕ**

* В виде единой матрицы графиков построить графики распределения времени доставки для каждого номера заказа. Каждый отдельный график в матрице должен иметь свой цвет и на каждом из графиков максимум необходимо выделить любым цветом, отличным от основного цвета выбранного графика.

1. **Кластеризация**

* Любым удобным методом кластеризуйте пользователей, на основании переменных "время заказа" и "сумма заказа"
* Обоснуйте число выбранных кластеров и присвойте их значения пользователям в отдельном столбце

1. **АБ**

* Выберите два любых кластера пользователей и проведите их сравнение
* **ЗАДАНИЕ ПОСЛОЖНЕЕ – НЕОБЯЗАТЕЛЬНОЕ**Выберите три любых кластера и проведите их сравнение