

## Лабораторная работа №1. Python и Pandas. Повторение

**Цель работы:** закрепить навыки работы с табличными данными с использованием библиотеки Pandas.

**Задания:** приведенные ниже задания были взяты из соревнований компании Яндекс по аналитике данных.

**Задание 1.** Финикийцы были известны своей активной торговлей по всему Средиземноморью. Но удивительно даже не это, а то, что уже в X веке нашей эры финикийские торговцы документировали все свои сделки, ведя учет о совершенных операциях и отчитываясь перед местными властями. Археологи обнаружили многочисленные свидетельства таких записей в разных частях Средиземноморья — в Марокко, Ливане, Испании, на Сицилии и в других местах. Мы собрали все доступные данные об этих операциях и преобразовали их в таблицу для дальнейшего анализа.

Таблица transactions (торговые операции):

- logid — идентификатор события
- date — время события
- countertype — тип события, может быть "sell" (продажа) или "buy" (покупка)
- citizenid — идентификатор гражданина, который совершил операцию
- marketid — идентификатор рынка, на котором была совершена операция
- merchantid — идентификатор торговца, с которым велась операция
- cost — деньги, затраченные на операцию
- model\_cost — прогнозная оценка стоимости операции
- fraudbits — фродовая операция или нет (0 — нет, 1 — да))

Таблица merchants (купцы):

- merchantid — идентификатор торговца, с которым велась операция
  - guild — гильдия торговца
  - coef — коэффициент амнистии торговца
1. Определите гильдию торговцев с наибольшей эффективностью за этот день. Эффективность определяется как произведение:

Соотношения событий продаж к покупкам  $\frac{sell}{buy}$ .

Средней амнистии на операцию. Амнистия считается  $\frac{|cost - model\_cost|}{coef}$ .

2. Эффективность этой гильдии.
3. Определите час, в который представители этой гильдии совершили наибольшее количество продаж.

Учитываются только данные, которые не отмечены как фродовые (fraudbits = 0).

Файлы с данными: merchants.csv, transactions.csv.

### Формат вывода:

Эффективность округлите с точностью до сотых, пример ответа:

Гильдия Пекарей 12.34 15

**Задание 2.** Цифровой художник решил монетизировать своё творчество. Он открыл свою галерею цифрового искусства в формате мобильного приложения. Галерея монетизируется через покупку пожизненной платной подписки — она позволяет получить изображение в высоком качестве. Подписку можно купить лишь один раз и нельзя отменить.

Художник считает хорошей метрикой: число активированных платных подписок за первые 90 дней с начала использования приложения (день начала использования будет первым и до 90-го дня включительно).

Художник передал вам данные по всем пользователям с момента запуска сервиса в CSV-файле с полями: ID пользователя, дата использования сервиса, бинарная переменная с фактом оплаты. Файл с данными: `startups_users_visits.csv`.

Дату начала использования считайте по первому визиту в логах. При этом группируйте пользователей по месяцу начала использования сервиса.

На выходе художник ожидает от вас CSV-файл (разделитель: запятая), где не будет строки заголовка, и для каждого месяца будет строка: первый день месяца, количество оплат от пользователей в течение 90 дней с момента первого визита (первый визит которых пришёлся на этот месяц). Месяцы должны быть упорядочены по возрастанию даты. Последний месяц, который должен быть в итоговом CSV-файле: июнь 2023 г.

**Задание 3.** Куратор сверхмасштабной международной выставки современного искусства “M-art: dialogue” выяснил, что семь художников, которые готовили работы на экспозицию, поступили нечестно. Они не сами писали картины, за них работала нейросеть. Этот факт обесценивает работы художников и выставку в целом. Если арт-критикам и публике станет известно об обмане, репутация команды и куратора будет разрушена.

Всего в датасете 199 художников. Известно, что работы семи из них сделаны с помощью ИИ. Куратор выяснил, как художники перевозили свои работы. Настоящие картины перевозились без аномальных историй, а работы, сгенерированные нейросетью демонстрировали очень странные свойства.

В датасете 9 разных фич:

`order` — идентификатор поездки  
`driver` — идентификатор художника  
`user` — идентификатор компаньона художника  
`zone` — зона поездки  
`coupon` — краткое название промокода поездки  
`currency` — валюта поездки  
`fact_km` — фактическое расстояние поездки  
`plan_km` — ожидаемое расстояние поездки  
`plan_time` — ожидаемое время поездки

Ваша задача: найти по этим фичам семерых нечестных художников.

**Решение будет зачтено, если среди 7 предоставленных id-шников хотя бы 6 будут правильными.**

Данные находятся в файле `data_fix.txt`.

## Оформление отчетов

Отчетом по данной работе является Python-ноутбук с ходом выполнения заданий лабораторной. В ноутбуке должны быть выведены ответы на задания (во

втором задании можно напечатать в ячейке содержимое датафрейма).

После защиты лабораторной работы отчет необходимо загрузить по ссылке: [https://vyatsu-my.sharepoint.com/:f:/g/personal/usr09019\\_vyatsu\\_ru/Ehnh-bqe0kFAk-DG\\_WSz8kMBQzGHOOr1T1C\\_btAVpsQjnGw](https://vyatsu-my.sharepoint.com/:f:/g/personal/usr09019_vyatsu_ru/Ehnh-bqe0kFAk-DG_WSz8kMBQzGHOOr1T1C_btAVpsQjnGw)