

Домашнее задание

Дисциплина	Python для инженерии данных
Тема	Тема 7. Продвинутая функциональность Pandas
Форма проверки	Самопроверка. Студент выполняет задание и самостоятельно проверяет его
Имя преподавателя	Дмитрий Клабуков
Время выполнения	2 часа
Цель задания	Научиться оптимизировать данные
Инструменты для выполнения ДЗ	Jupyter Notebook или Google Colab
Правила приёма работы	Ссылку на задание, выполненное в Google Colab или GitHub (если вы использовали Jupyter Notebook), прикрепите в личном кабинете. Важно: убедитесь, что по ссылке есть доступ в Google Colab для другого аккаунта
Критерии оценки	Задание считается выполненным, если: <ul style="list-style-type: none">- прикреплена ссылка на файл с выполненным заданием,- доступ к файлу открыт,- код даёт правильный ответ к задаче. Задание не выполнено, если: <ul style="list-style-type: none">- файл с заданием не прикреплен или отсутствует доступ по ссылке.- код выдаёт ошибку или даёт неправильный ответ
Дедлайн	7 дней с даты проведения соответствующего вебинара

Описание задания

Перед выполнением задания установите Jupyter Notebook или используйте Google Colab.

Скачайте [базу данных по пользователям](#).

Сущности базы данных:

- contact_preference — справочник предпочтительного способа связи;
- customer_preferenced_contact_type — таблица связи пользователей (предпочтительный способ связи);

- customer — справочник пользователей с основной информацией по ним;
- customers_segmetns — таблица связей пользователей и сегментов;
- location — справочник локаций;
- product_feedback — отзывы по продуктам;
- product — справочник продуктов с именем, категорией, ценой;
- purchase — таблица покупок;
- segment — справочник сегментов с информацией о названии сегмента и скидки;
- service_feedback — отзывы о сервисе.

Задача 1: собрать всю информацию по каждому пользователю в отдельной витрине, а также на основе имеющихся данных посчитать:

- количество покупок,
- количество потраченных средств,
- средний рейтинг по продукту,
- наивысший продукт по рейтингу,
- самый популярный тип продукта,
- количество отзывов.

Задача 2: написать функцию, которая считает долю трат пользователя по разным категориям от общих трат пользователя. Категории: Movies, Games, Books.

Чек-лист самопроверки

Критерии выполнения задания	Отметка о выполнении
Установлен Jupyter Notebook или используется Google Colab	
Создан профиль на GitHub , если используете Jupyter Notebook	
Вся информация по пользователю собрана в отдельной витрине	
Задача 1. Для пользователя вычислены: <ul style="list-style-type: none"> • количество покупок, • количество потраченных средств, • средний рейтинг по продукту, 	

<ul style="list-style-type: none"> • наивысший продукт по рейтингу, • самый популярный тип продукта, • количество отзывов. 	
<p>Задача 2.</p> <p>Функция считает долю трат пользователя по категориям Movies, Games, Books от общих трат пользователя</p>	
<p>На учебной платформе прикреплена ссылка на выполненное задание в Google Colab или GitHub</p>	
<p>Есть доступ по ссылке, если используете Google Colab</p>	