

Лабораторная работа по курсу баз данных

Составители: Michael, Miguel, Marat (MMM)

28 апреля 2025 г.

Инструкции:

- Если работа была списана и/или сделана LLM, то за работу ставится **-1 балл** (вычитается из проекта/зачета).
- Все задания должны быть выполнены **только с использованием SQL**.
- Загрузите результаты вашего анализа и коды в readme-файл, включая комментарии, объясняющие ваши шаги и выводы. Решение необходимо залить на Гитлаб и поставить меня и ассистентов в ревьюеров.
- Крайний срок сдачи: **08.05.2025 8:08**.
- Ссылка на датасет
- За решение не своего варианта ваши баллы поделятся на $\text{rand}(2, 3)$
- Вариант = $(\lfloor \text{Номер первой буквы фамилии} \rfloor + \lfloor \text{Номер первой буквы имени} \rfloor + \lfloor \text{Номер второй буквы логина Telegram} \rfloor) \bmod 7 + 1$
p.s. латиница

Вариант 1

1. Сравните заказы из городов Сан-Паулу, Рио-де-Жанейро, Белу-Оризонти, Бразилиа, Салвадор с остальными городами и выясните, насколько сильно отличаются категории товаров.
2. Определите, насколько сильно отличается стоимость доставки в зависимости от региона. Выдвиньте гипотезы и проверьте их. Объясните ваше решение.

3. Найдите селлеров с самыми высокими скидками и подумайте, насколько эффективна их стратегия.
4. Исследуйте распределение количества заказов по способам оплаты в ТОП-5 городах.

Вариант 2

1. Проанализируйте количество фотографий товара в зависимости от стоимости для разных категорий товаров и объясните основные причины.
2. Проанализируйте в городе Рио-де-Жанейро распределение товаров электроника, косметика, игрушки по месяцам и объясните, почему распределение именно такое.
3. Выведите id 10% пользователей из штата Рио-де-Жанейро, у которых суммарная стоимость всех заказов (средняя стоимость заказа, количество заказов) с 2017-05-07 по 2018-06-14 максимальна.
4. Рассчитайте долю выручки ТОП-10% самых “плотных” товаров относительно всех товаров в штате, где столица Салвадор (подсказка: плотность = масса / объем).

Вариант 3

1. Определите, на какие категории товаров в штате, где столица Бразилиа, пользователи чаще всего оставляют комментарии и подумайте, почему они это делают.
2. Исследуйте скорость заполнения пользователями формы о качестве заказа.
3. Рассчитайте средний чек клиента для каждого штата.
4. Найдите категории товаров, которые чаще всего попадают в отмененные заказы.

Вариант 4

1. Определите области, где клиенты тратят больше всего на заказ, и сделайте выводы.

2. Сравните заказы из ТОП-5 крупнейших городов (Сантус, Манаус, Куритиба, Гояния, Кампинас) с другими регионами по количеству возвращенных товаров.
3. Проанализируйте эффективность маркетинговых кампаний в различных сезонах.
4. Определите долю пользователей, которые после первого заказа в течение 60 дней совершили еще один заказ.

Вариант 5

1. Исследуйте распределение количества заказов по способам оплаты, а также в ТОП-5 регионах.
2. Рассчитайте долю товаров каждой категории по количеству заказов в штате со столицей Порту-Алегри.
3. Вычислите долю выручки ТОП-10% самых “плотных” товаров относительно всех товаров в штате Сан-Паулу (подсказка: плотность = масса / объем).

Вариант 6

1. Проанализируйте для города Белу-Оризонти распределение товаров электроника, мебель, игрушки по месяцам и объясните с чем это связано.
2. Определите категории товаров, которые чаще всего попадают в отмененные заказы штата Рио-де-Жанейро.
3. Определите средний чек клиента для штатов Сан-Паулу и Пернамбуку.
4. Исследуйте долю пользователей, которые с момента первого заказа в течение 30 дней совершили еще один заказ.

Вариант 7

1. Проанализируйте, как распределяются заказы по способам оплаты в ТОП-5 городах по количеству заказов.

2. Рассчитайте, какую долю по количеству заказов занимает каждая категория в штате Риу-Гранди-ду-Норти.
3. Вычислите долю выручки ТОП-10% самых “плотных” товаров относительно всех товаров в штате Парана (подсказка: плотность = масса / объем).