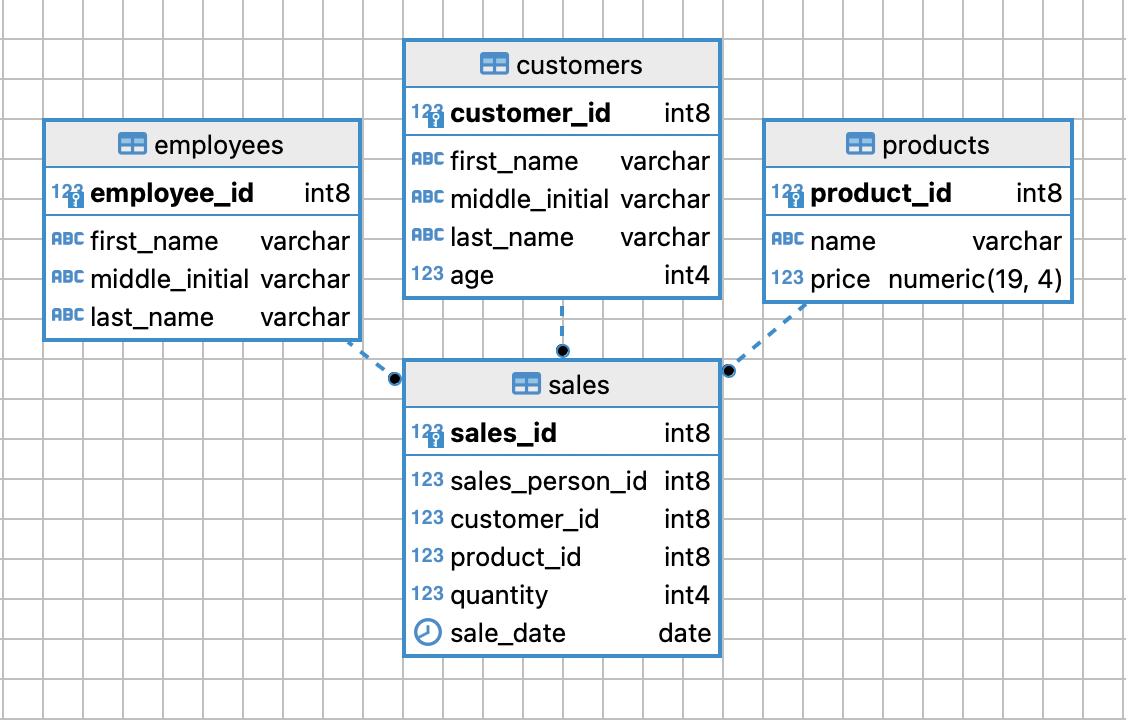
Проект Продажи

Работа с базой данных

Когда аналитик работает в компании, ему удобнее получать данные напрямую из источников. Особенно это важно, когда данные должны быть живыми и обновляться практически сразу. Получать такие данные удобно напрямую из базы данных. Аналитик может выполнять запросы напрямую, агрегируя данные и собирая нужные таблицы.

Вам был предоставлен доступ к основной базе данных торговой площадки. На этом шаге вы подключитесь к ней и сможете читать из нее данные.

Описание базы данных



База данных состоит из четырех таблиц.

*customers* — таблица покупателей. Колонки:

* customer\_id — ID покупателя
* first\_name — имя
* middle\_initial — инициал отчества
* last\_name — фамилия
* age - возраст покупателя

*employees* — таблица сотрудников отдела продаж:

* employee\_id — ID сотрудника
* first\_name — имя
* middle\_initial — инициал отчества
* last\_name — фамилия

*products* — таблица товаров:

* product\_id — ID товара
* name — название
* price — цена

*sales* — таблица с данными о проданных товарах:

* sales\_id
* sales\_person\_id — ID сотрудника, продавшего товар
* customer\_id — ID покупателя
* product\_id — ID товара
* quantity — количество
* sale\_date — дата продажи

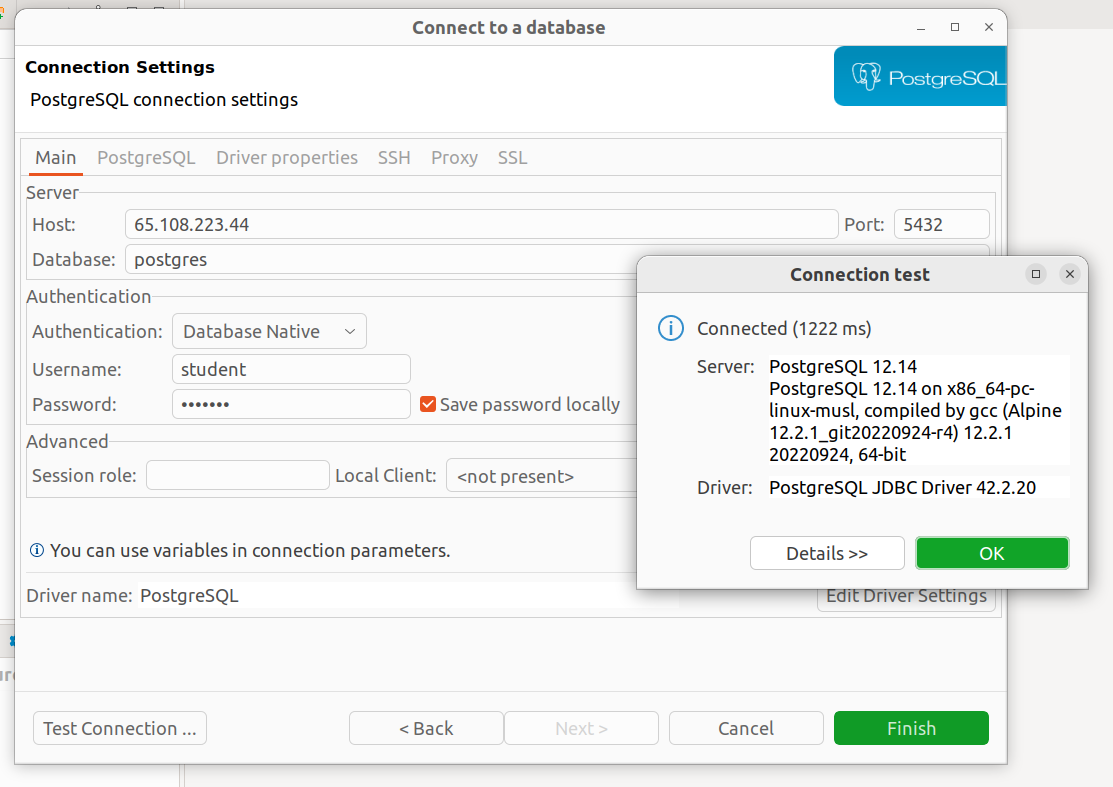
Ссылки

* [DBeaver](https://dbeaver.io/download/) — инструмент для работы с базами данных
* [Инструкция по подключению к базе данных](https://dbeaver.com/docs/wiki/Create-Connection/)

Задачи

* Скачайте [DBeaver](https://dbeaver.io/download/" \t "_blank) — программу для работы с базой данных. Выберите PostgreSQL и по [инструкции](https://dbeaver.com/docs/wiki/Create-Connection/) создайте новое подключение. Используйте следующие данные:
  + Host: 65.108.223.44
  + Database: salesdb
  + Username: student
  + Password: student
  + Port: 5432

Нажмите *Test connection* и проверьте, что соединение сработало:



* Подключитесь к базе данных
* Напишите запрос, который считает общее количество покупателей из таблицы *customers*. Назовите колонку *customers\_count*
* Экспортируйте результат в CSV-файл под названием *customers\_count.csv*. Результат в файле должен получиться примерно такой:

"customers\_count"

1234

* Создайте файл *queries.sql* и запишите туда запрос. С помощью комментария опишите, что выполняет данный запрос

Анализ отдела продаж

Компания хочет проанализировать работу продавцов. Чтобы в этом помочь, предоставим несколько отчетов о прошедшей работе.

Первый отчет о десятке лучших продавцов. Таблица состоит из трех колонок - данных о продавце, суммарной выручке с проданных товаров и количестве проведенных сделок, и отсортирована по убыванию выручки:

* seller — имя и фамилия продавца
* operations - количество проведенных сделок
* income — суммарная выручка продавца за все время

Пример таблицы:

|seller |operations |income |

|-----------------|-----------|---------------|

|Vasya Pupkin |100 |100500 |

|Petya Ivanov |80 |54320 |

|Anna Dmitrieva |50 |3573 |

// остальные записи

Второй отчет содержит информацию о продавцах, чья средняя выручка за сделку меньше средней выручки за сделку по всем продавцам. Таблица отсортирована по выручке по возрастанию.

* seller — имя и фамилия продавца
* average\_income — средняя выручка продавца за сделку с округлением до целого

| seller | average\_income |

|-----------------|----------------|

|Anthony Smith |25 |

|John Cristofer |145 |

|Harry Potter |2503 |

// остальные записи

Третий отчет содержит информацию о выручке по дням недели. Каждая запись содержит имя и фамилию продавца, день недели и суммарную выручку. Отсортируйте данные по порядковому номеру дня недели и seller

* seller — имя и фамилия продавца
* day\_of\_week — название дня недели на английском языке
* income — суммарная выручка продавца в определенный день недели, округленная до целого числа

|seller |day\_of\_week |income |

|-----------------|-------------|-------|

|Albert Rudkov |monday |9237 |

|Anton Borisov |monday |3184 |

|Albert Rudkov |tuesday |5905 |

|Anton Borisov |tuesday |5905 |

|Albert Rudkov |wednesday |1484 |

|Anton Borisov |wednesday |1484 |

|Albert Rudkov |thursday |2613 |

|Anton Borisov |thursday |2613 |

|Albert Rudkov |friday |5749 |

|Anton Borisov |friday |3393 |

|Albert Rudkov |saturday |5905 |

|Anton Borisov |saturday |1935 |

|Albert Rudkov |sunday |6298 |

|Anton Borisov |sunday |3232 |

// далее следуют еще записи

Ссылки

* [TO\_CHAR()](https://www.postgresql.org/docs/current/functions-formatting.html#id-1.5.8.14.4.2.2.1.1.1.1) - функция для преобразования значений даты/времени в строку
* [EXTRACT()](https://www.postgresql.org/docs/12/functions-datetime.html#FUNCTIONS-DATETIME-EXTRACT) — функция, которая позволяет из значений даты/времени получать определенные поля
* [CONCAT()](https://postgrespro.ru/docs/postgrespro/10/functions-string#id-1.5.8.9.7.2.2.4.1.1) — функция для "склейки" двух строк или полей. Также в склейке значений может помочь оператор ||
* [FLOOR()](https://www.postgresqltutorial.com/postgresql-math-functions/postgresql-floor/)
* [TRIM()](https://www.postgresql.org/docs/current/functions-string.html#id-1.5.8.10.5.2.2.20.1.1.1) — функция пригодится, чтобы удалить лишние пробелы из строки

Задание

1. Подготовьте в файл *top\_10\_total\_income.csv* отчет с продавцами у которых наибольшая выручка
2. Подготовьте в файл *lowest\_average\_income.csv* отчет с продавцами, чья выручка ниже средней выручки всех продавцов
3. Подготовьте в файл *day\_of\_the\_week\_income.csv* отчет с данными по выручке по каждому продавцу и дню недели
4. Добавьте запросы, при помощи которых были подготовлены отчеты, в файл *queries.sql*. Не забудьте оставить комментарии с описанием, что делает каждый запрос
5. Загрузите все отчеты и обновленный файл с запросами на Github

Все данные нужно округлить, избавившись от чисел после запятой, т.е. округлить в меньшую сторону.

Анализ покупателей

Поможем компании проанализировать данные о покупателях и покупках. Для того, чтобы оказать помощь в этом деле, вам необходимо предоставить следующие данные:

Первый отчет - количество покупателей в разных возрастных группах: 16-25, 26-40 и 40+. Итоговая таблица должна быть отсортирована по возрастным группам и содержать следующие поля:

* age\_category - возрастная группа
* age\_count - количество человек в группе

|age\_category |age\_count |

|-----------------------|-------------|

|16-25 |3 |

|26-40 |123 |

|40+ |7 |

Во втором отчете предоставьте данные по количеству уникальных покупателей и выручке, которую они принесли. Сгруппируйте данные по дате, которая представлена в числовом виде ГОД-МЕСЯЦ. Итоговая таблица должна быть отсортирована по дате по возрастанию и содержать следующие поля:

* date - дата в указанном формате
* total\_customers - количество покупателей
* income - принесенная выручка

|selling\_month |total\_customers |income |

|--------------|----------------|--------|

|2003-01 |2000 |256789 |

|2003-02 |768 |34567 |

|2003-03 |42 |3457 |

// остальные записи

Третий отчет следует составить о покупателях, первая покупка которых была в ходе проведения акций (акционные товары отпускали со стоимостью равной 0). Итоговая таблица должна быть отсортирована по id покупателя. Таблица состоит из следующих полей:

* customer - имя и фамилия покупателя
* sale\_date - дата покупки
* seller - имя и фамилия продавца

|customer |sale\_date |seller |

|----------------|-------------------|----------------|

|Krista Gill |2012-12-21 |Dirk Stringer |

|Mario Rai |1999-10-09 |Marjorie Green |

// остальные записи

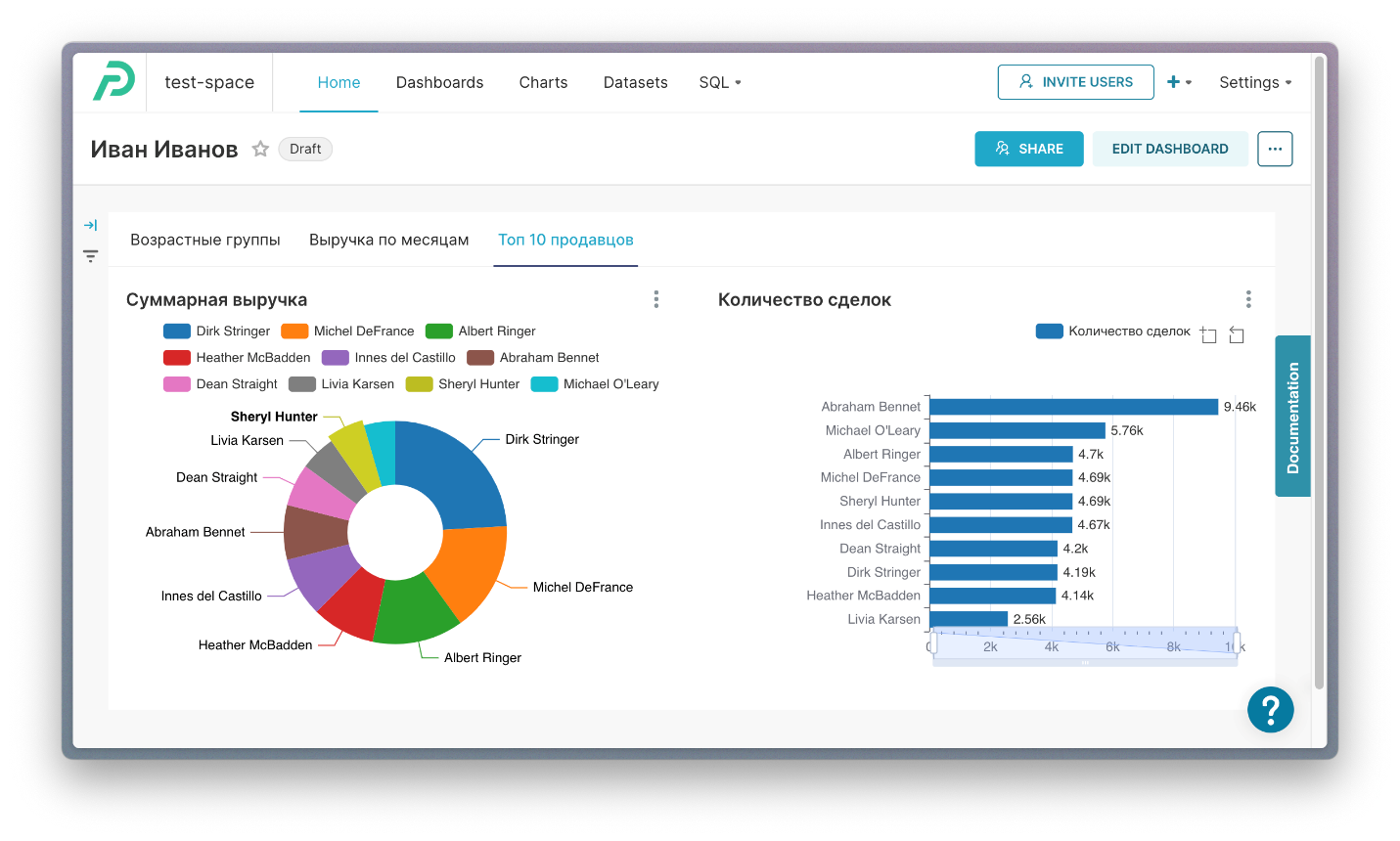
В таблице

Задачи

1. Подготовьте в файл *age\_groups.csv* с возрастными группами покупателей
2. Подготовьте в файл *customers\_by\_month.csv* с количеством покупателей и выручкой по месяцам
3. Подготовьте в файл *special\_offer.csv* с покупателями первая покупка которых пришлась на время проведения специальных акций
4. Добавьте запросы, при помощи которых были подготовлены отчеты, в файл *queries.sql*. Не забудьте оставить комментарии с описанием, что делает каждый запрос
5. Загрузите все отчеты и обновленный файл с запросами на Github

Подсказки

* Не забудьте про округление
* Дашборды
* Результаты проведенного исследования продавцов и покупателей следует представить в виде дашборда, который наилучшим образом отразит полученные результаты.



Презентация

Полученные данные необходимо представить в виде презентации с анализом полученных результатов. Презентация должна быть интересной, наглядной и понятной для аудитории, не знакомой с дата-аналитикой. Она должна содержать ссылку на итоговый дашборд.

Ссылки

* [Инструмент для создания презентаций Google Slides](https://docs.google.com/presentation/)

Задачи

* Выберите наиболее важные и релевантные графики и диаграммы из проекта, ключевые метрики и показатели, а также интересные закономерности
* Сохраните графики и диаграммы в формате изображений с помощью Preset
* Создайте новую презентацию в Google Slides. Выберите оформление слайдов по своему усмотрению
* Вставьте изображения на отдельные слайды. Распределите их таким образом, чтобы они логично следовали друг за другом
* Добавьте к каждому слайду с графиком или диаграммой краткое пояснение, в котором укажите, что показывает график или диаграмма, какие выводы можно сделать из него
* Добавьте ссылку на дашборд в презентацию
* Выгрузите полученную презентацию в PDF
* Загрузите ее на Github с именем *presentation.pdf*