Департамент образования и науки города Москвы

Государственное автономное образовательное учреждение высшего образования города Москвы

«Московский городской педагогический университет»

Институт цифрового образования

Департамент информатики, управления и технологий

ДИСЦИПЛИНА:

Проектный практикум по разработке ETL-решений

**Лабораторная работа 2**

**Динамические соединения с базами данных**

Выполнил(а): st\_98 группа: АДЭУ-211

Преподаватель:

Москва

2025

**Цель работы:** получить практические навыки создания ETL-процесса для загрузки данных из CSV-файла в базу данных MySQL с использованием Pentaho Data Integration.

**Задачи:**

− Создать динамические подключения к различным источникам данных.

− Разработать процесс выявления и обработки дублирующихся записей. − Реализовать механизм объединения данных в единое хранилище.

− Настроить обработку ошибок при выполнении трансформации. Программное обеспечение:

- Pentaho Data Integration 9.4. - MySQL или PostgreSQL. - CSV или Excel файлы с тестовыми данными.

**Вариант 11. Фильтр по сумме: Sales > 100. Статистика по категориям. Анализ клиентов**

На рисунке 1 продемонстрирована схема job

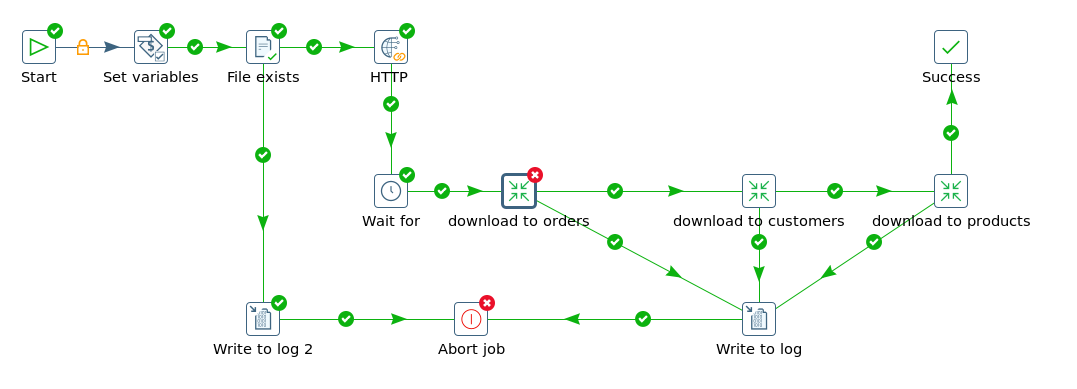


Рисунок 1. Job

Далее на рисунке 2 показана трансформация Orders

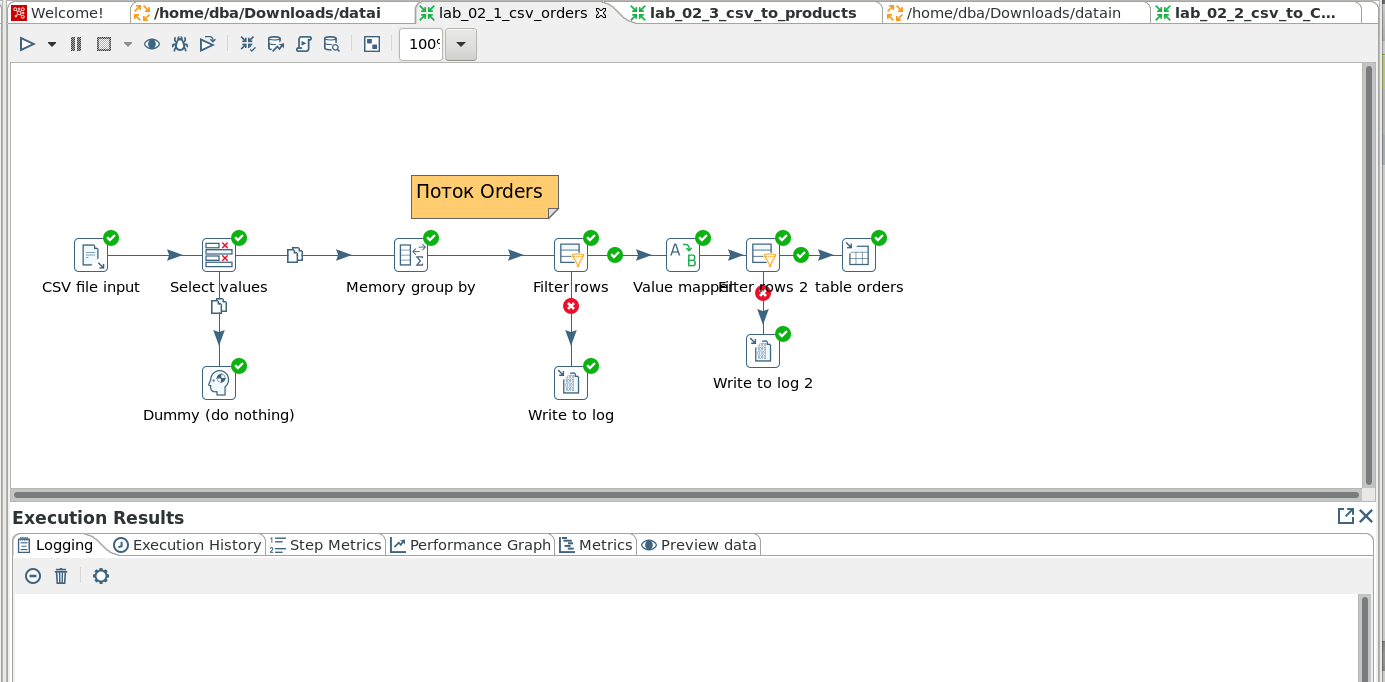


Рисунок 2. Трансформация Orders

В задании требовалось сделать фильтрацию про продажам, где они должны быть больше 50, на рисунке 3 это показано.

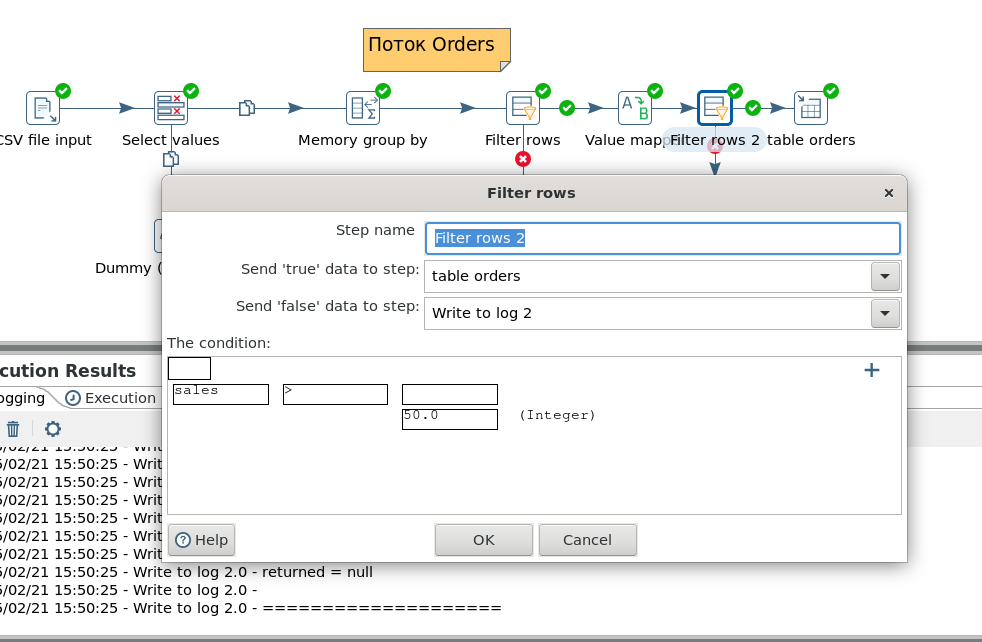
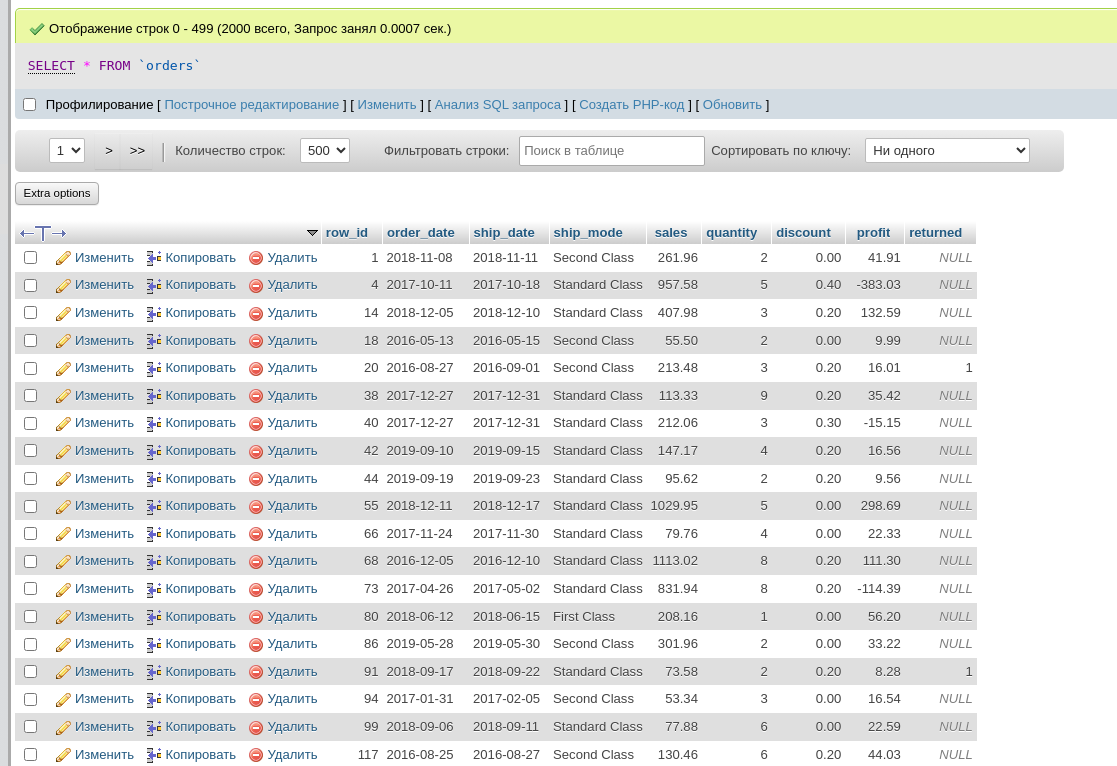


Рисунок 3. Фильтр продажи больше 50

На рисунке 4 показано, что таблица orders успешно загрузилось в базу данных с фильтрацией.

Рисунок 4. Таблица orders загрузилась в базу данных

На рисунке 5 показана трансформация Product, которая успешно отработала

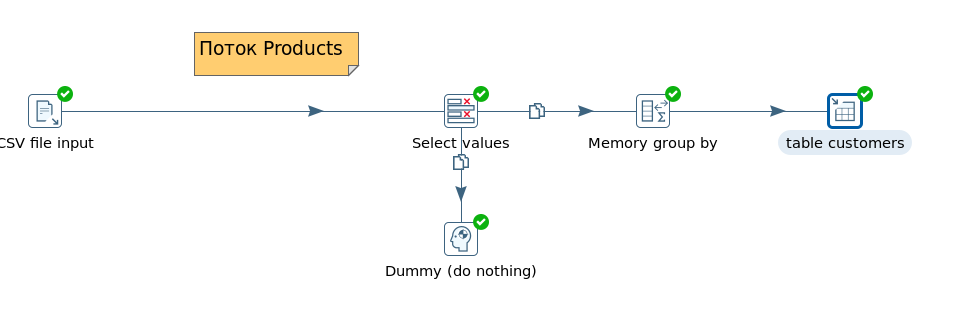


Рисунок 5. Трансформация Product

На рисунке 6 продемонстрировано, что таблица Product загрузилась в PhpMyAdmin

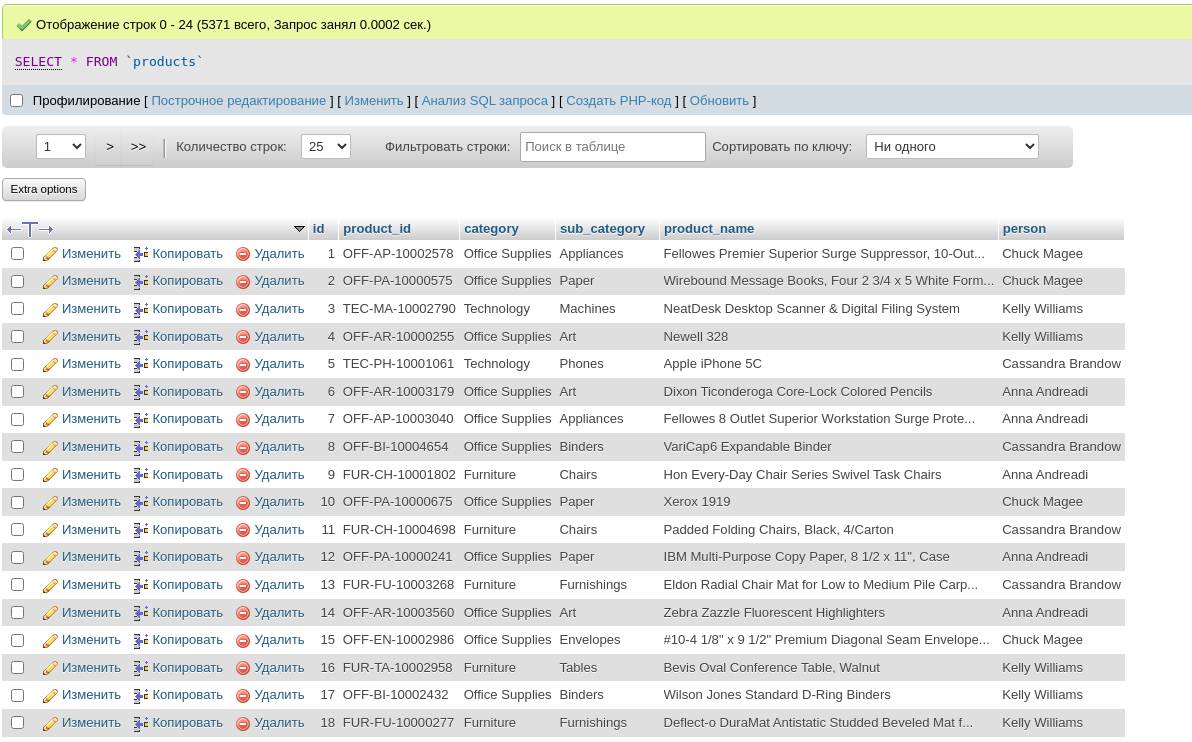


Рисунок 6. Таблица Product загрузилась в PhpMyAdmin

Затем нужно загрузить таблицу Customers. Трансформация представлена на рисунке 7.

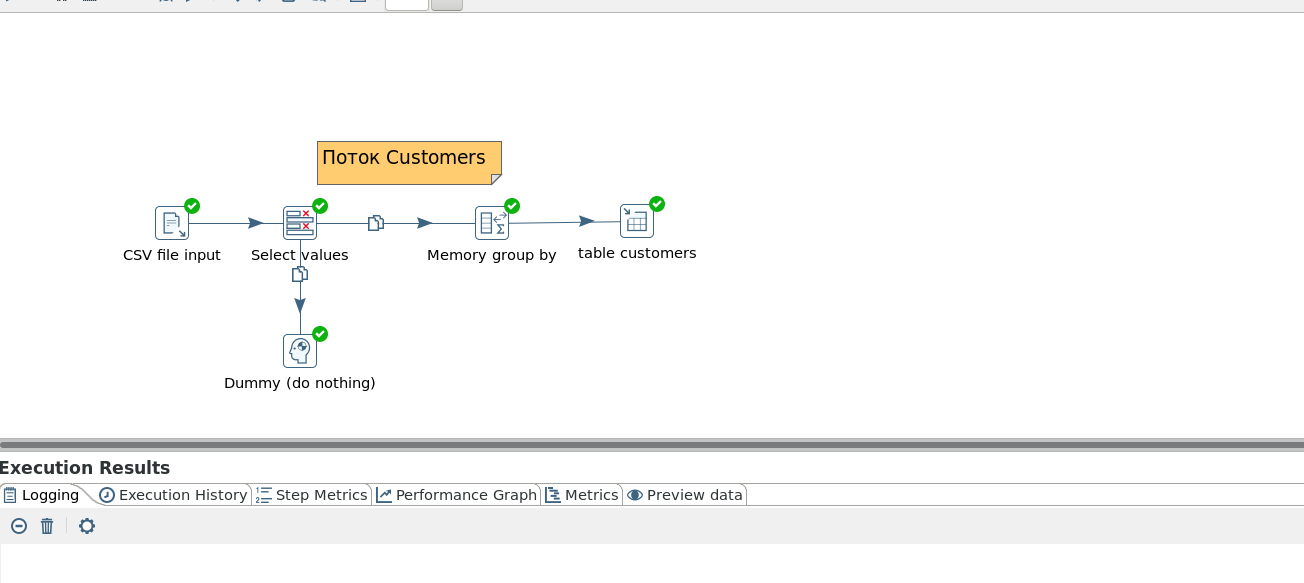


Рисунок 7. Трансформация Customers

На рисунке 8 показано, что трансформация успешно отработала и данные загрузились в таблицу Customers

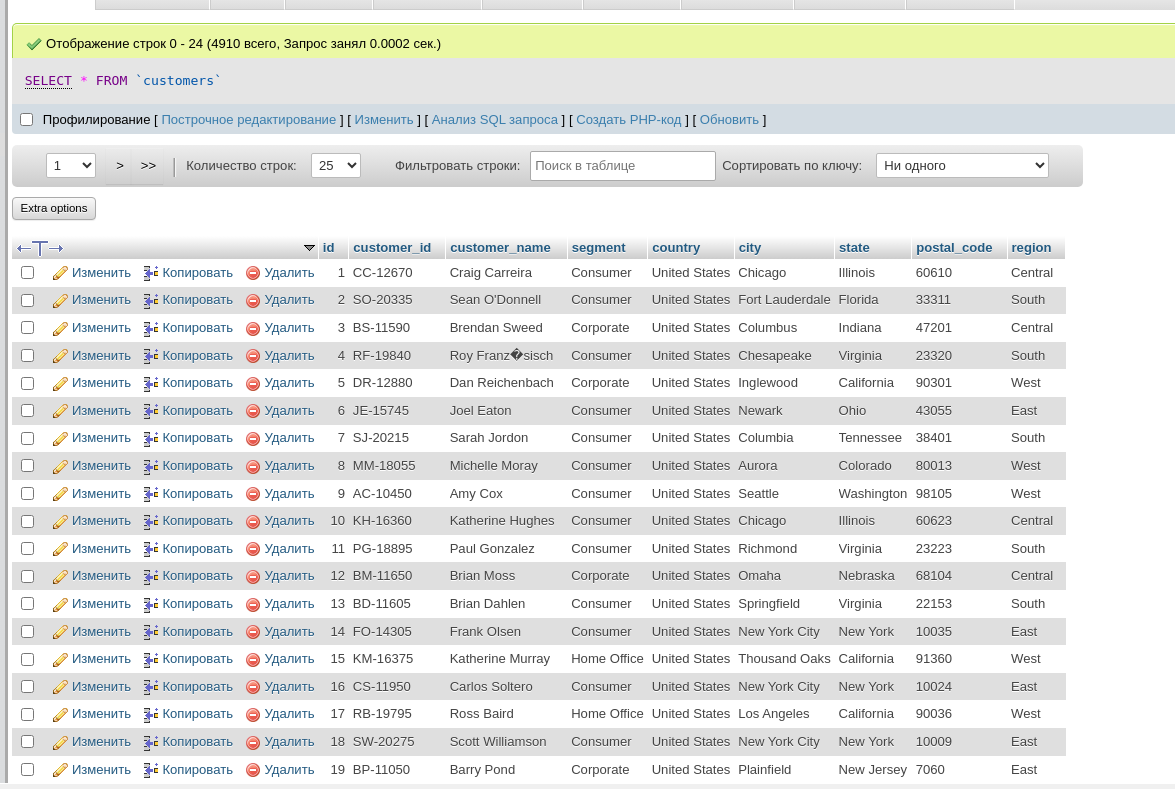


Рисунок 8. Данные Customers загрузились в базу данных

На рисунке 9 показана трансформация для статистики по категориям

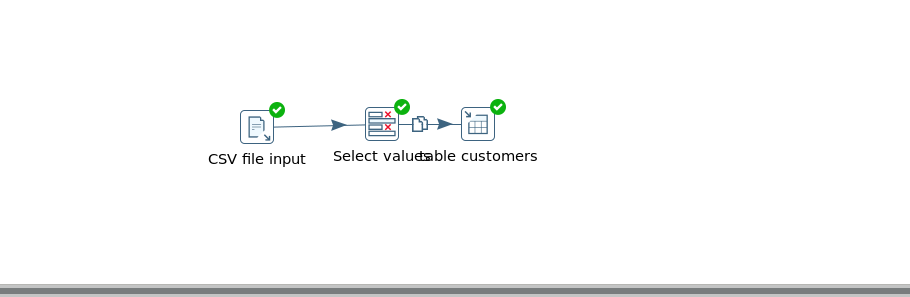


Рисунок 9. Трансформация для статистики по категориям

На рисунке 10 продемонстрировано, что таблица загрузилась, и можно увидеть такие столбцы, как категория, профит, количество и продажи.

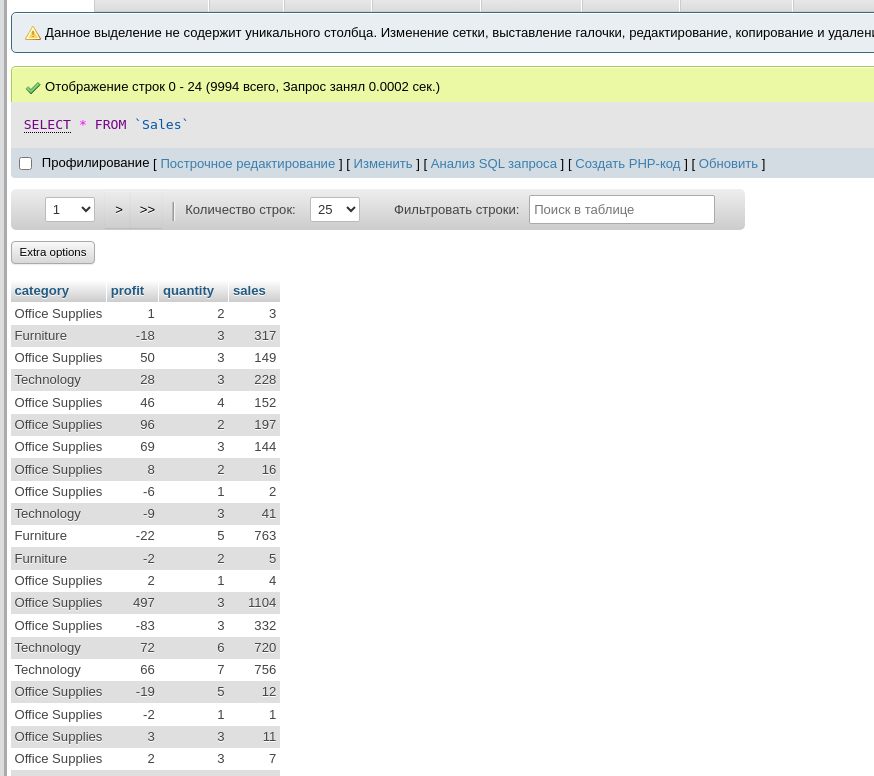


Рисунок 10. Таблица для статистики по категориям

На рисунке 11 показан sql-запрос, который выводит статистику профита, количества и продаж по каждой из категории

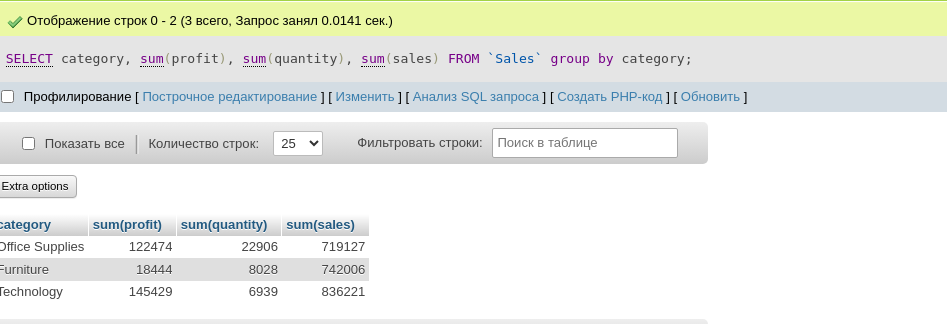


Рисунок 11. Статистика по категориям

Далее надо сделать анализ клиентов. На рисунке 12 показана трансформация.

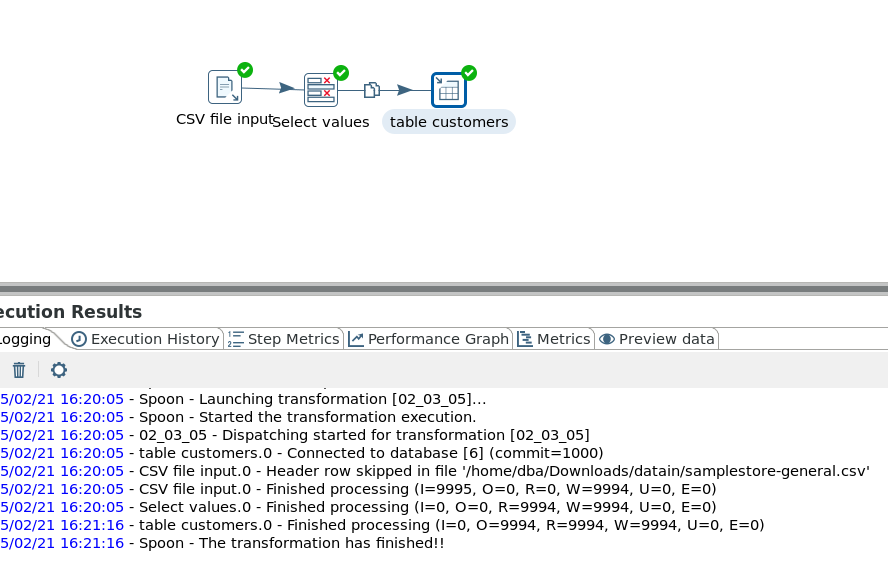


Рисунок 12. Трансформация для анализа клиентов

На рисунке 13 показано, что таблица успешно загрузилась в базу данных

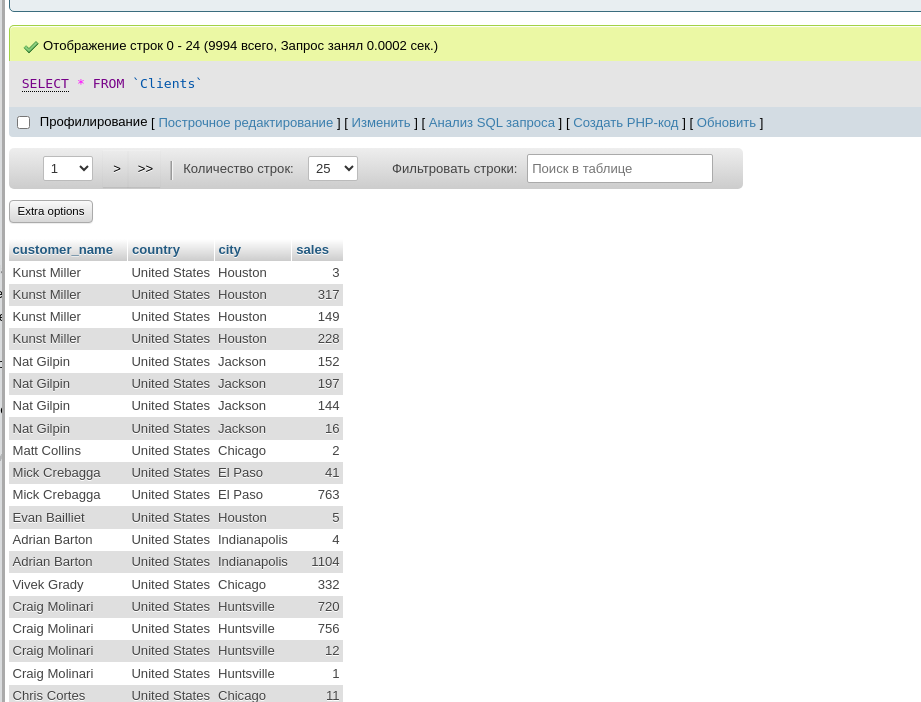


Рисунок 13. Таблица для анализа клиентов

На рисунке 14 показан анализ клиентов, сколько продаж из каждого города. Как можно увидеть, больше всего продаж из Нью-Йорка, Лос-Анджелеса и Сиэтла, так как это одни из самых крупных городов.



Рисунок 14. Анализ клиентов

**Выводы:**

1. Были созданы динамические подключения к различным источникам
2. Был разработан процесс выявления и обработки дублирующихся записей.
3. Был реализован механизм объединения данных в единое хранилище
4. Была настроена обработка ошибок при выполнении трансформации.