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 3.1**

**Интеграция данных из нескольких источников. Обработка и согласование данных из разных источников**

Выполнила: Шведова С.С., группа: АДЭУ-211

Преподаватель: Босенко Т.М.

Москва

2025

**Цель работы:** получить практические навыки интеграции, обработки и согласования данных из различных источников с использованием Python и его библиотек.

**Задачи:**

1. Необходимо нарисовать верхнеуровневую архитектуру аналитического решения в https://draw.io/
2. Изучить методы чтения данных из разных источников.
3. Освоить техники обработки и очистки данных.
4. Научиться согласовывать данные из разных источников.
5. Реализовать сохранение обработанных данных.

**Вариант 13. Интеграция данных о лояльности клиентов:**

- MySQL база данных бонусной программы.

- CSV файл с историей покупок.

- Excel файл с акциями и специальными предложениями.

Задача: проанализировать эффективность программы лояльности.

На рисунке 1 показано создание таблицы бонусной программы в MySQL.

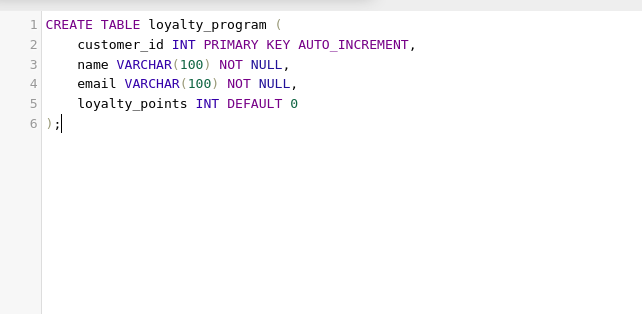


Рисунок 1. Создание таблицы бонусной программы

На рисунке 2 показана заполненность этой таблицы.

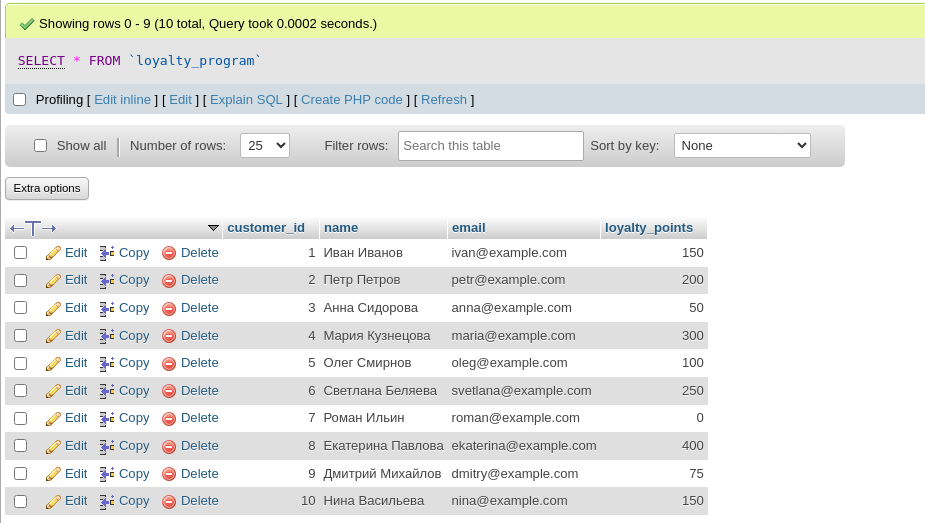


Рисунок 2. Заполненность таблицы

Верхеуровневая архитектура показана на рисунке 3.

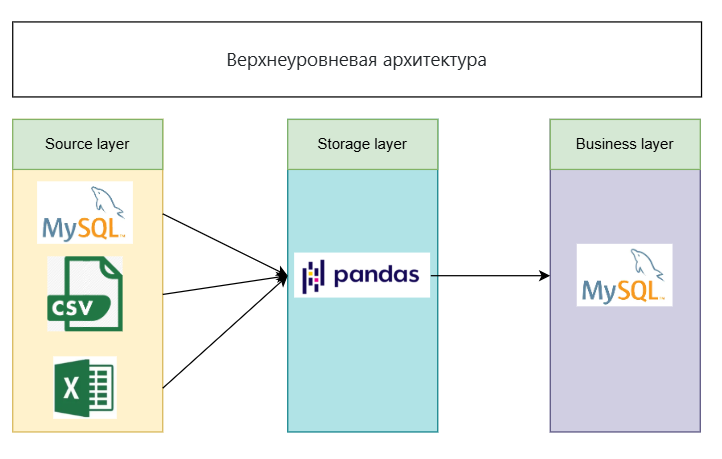


Рисунок 3. Верхнеуровневая архитектура

На рисунке 4 импорт нужных библиотек и настройка подключения к MySQL.

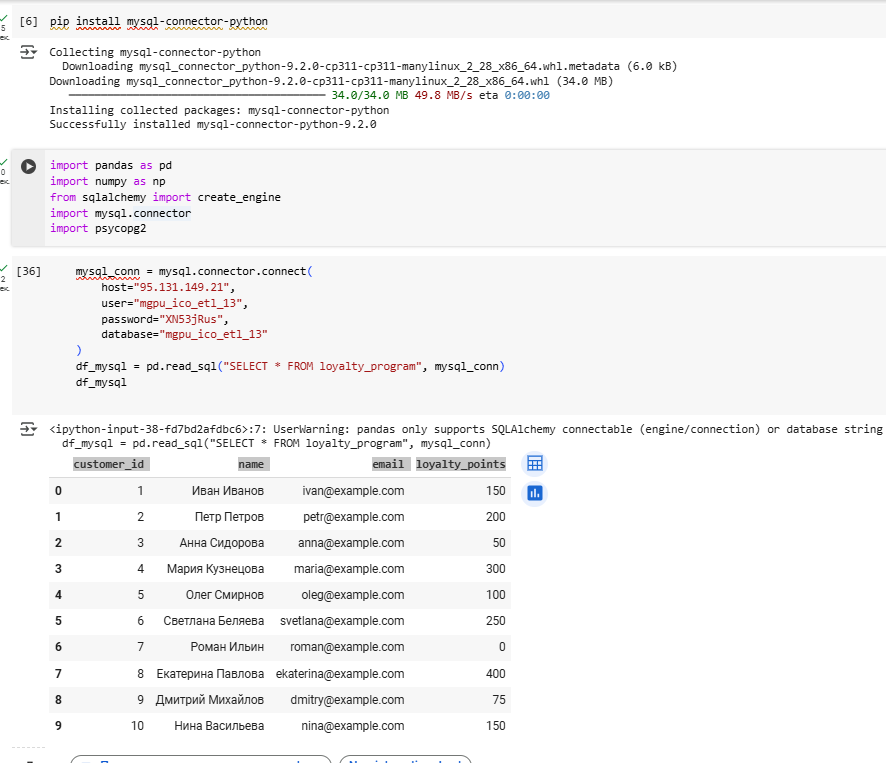


Рисунок 4. Настройка подключение к Mysql

На рисунке 5 показана очистка данных (дубликатов) таблицы бонусной программы, пропущенных значений и просмотр типов данных. А также те же самые действия проделаны с таблицей истории покупок.

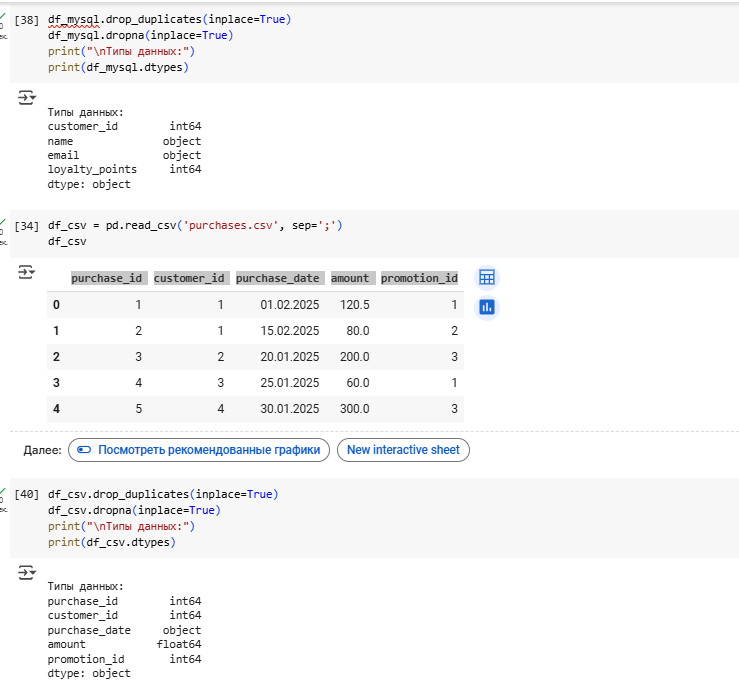


Рисунок 5. Просмотр типов данных, удаление дубликатов и пропущенных значений

Далее идет просмотр типов данных, удаление дубликатов и пропущенных значений файла с акциями и специальными предложениями и соединение трех таблиц (рисунок 6).

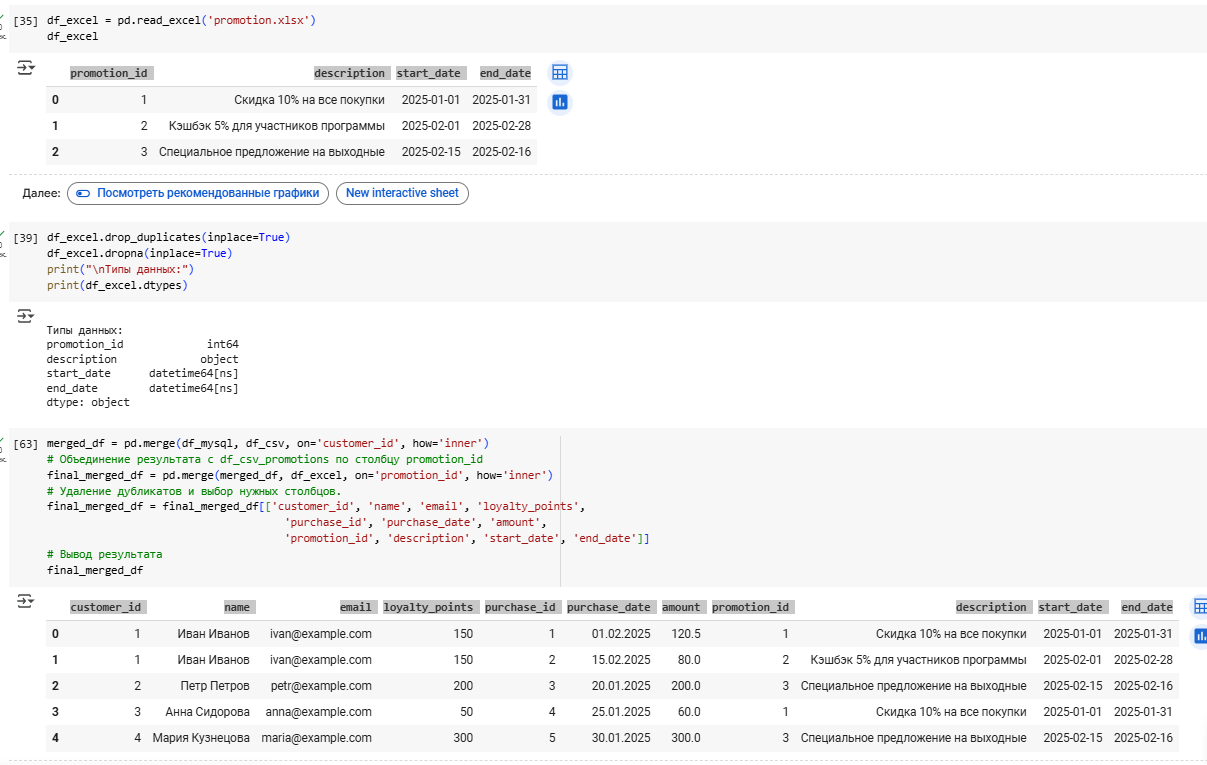


Рисунок 6. Соединение трех таблиц

На рисунке 7 показан анализ эффективности бонусной программы. Как можно заметить, средний уровень лояльности равен 175, клиентов 4 и покупок 5.

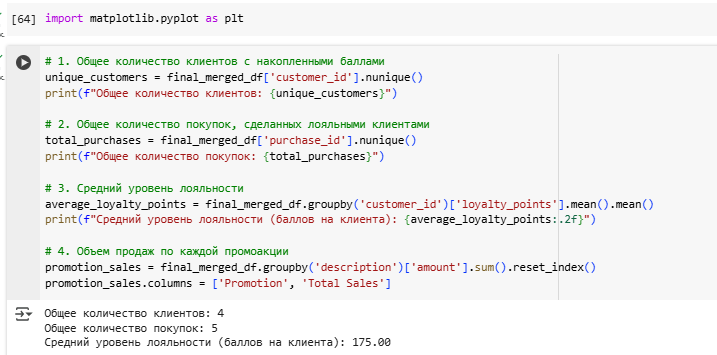


Рисунок 7. Анализ эффективности бонусной программы

С помощью библиотеки альтаир была построена столбчатая диаграмма объема продаж по промоакциям. Как можно заметить на рисунке 8, больше всего пользуются спросом специальные предложения на выходные.



Рисунок 8. Визуализация

На рисунке 9 показана загрузка в базу данных таблицы.



Рисунок 9. Загрузка в базу данных таблицы

На рисунке 10 видно, что таблица успешно загрузилась.

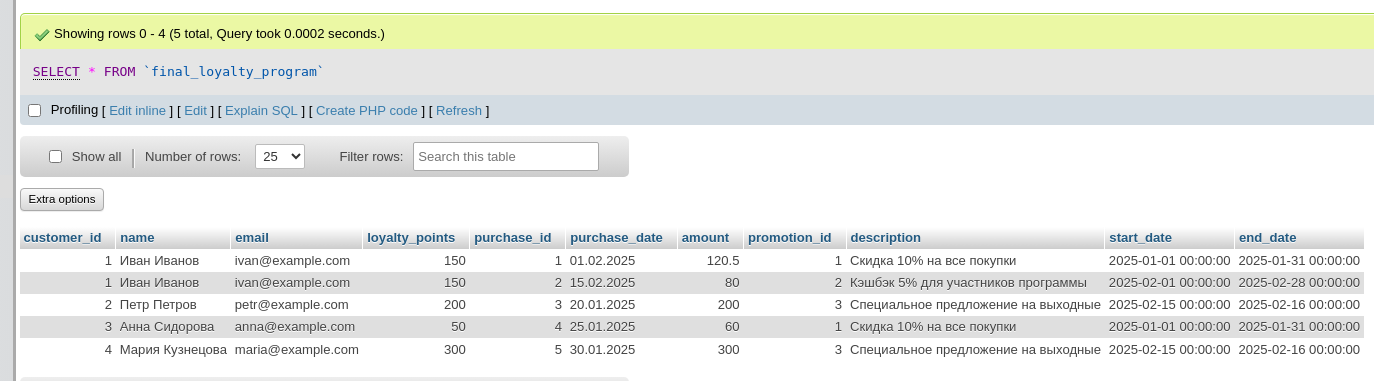


Рисунок 10. Таблица загрузилась в PhpMyAdmin

**Выводы:**

 Изучены методы чтения данных из разных источников.

 Освоены техники обработки и очистки данных.

 Согласованы данные из разных источников.

 Реализовано сохранение обработанных данных.