Лабораторная работа №4. Маскировка и анонимизация данных

Вариант №2

Выполнила студентка группы М3212

Авсюкевич Анастасия

**Задачи:**

1. Замаскировать поля с конфиденциальными данными.

2. Провести анонимизацию данных.

**Порядок выполнения работы:**

1. Установите расширение [PostgreSQL Anonymizer](https://postgresql-anonymizer.readthedocs.io/en/stable/INSTALL/)
2. Выберите поля, которые необходимо замаскировать и модифицируйте уже существующие таблицы или представления. [Динамическая маскировка данных](https://postgresql-anonymizer.readthedocs.io/en/stable/dynamic_masking/)
3. Выберите данные, которые можно анализировать, скрыв, обобщив конфиденциальные данные. Создайте три [Materialized Views](https://www.postgresql.org/docs/current/rules-materializedviews.html) используя:

[Generalization](https://postgresql-anonymizer.readthedocs.io/en/stable/generalization) – заменяет данные более широкими и менее точными значениями, диапазонами.

1. Используйте 2 [стратегии анонимизации](https://postgresql-anonymizer.readthedocs.io/en/stable/masking_functions/) из списка:

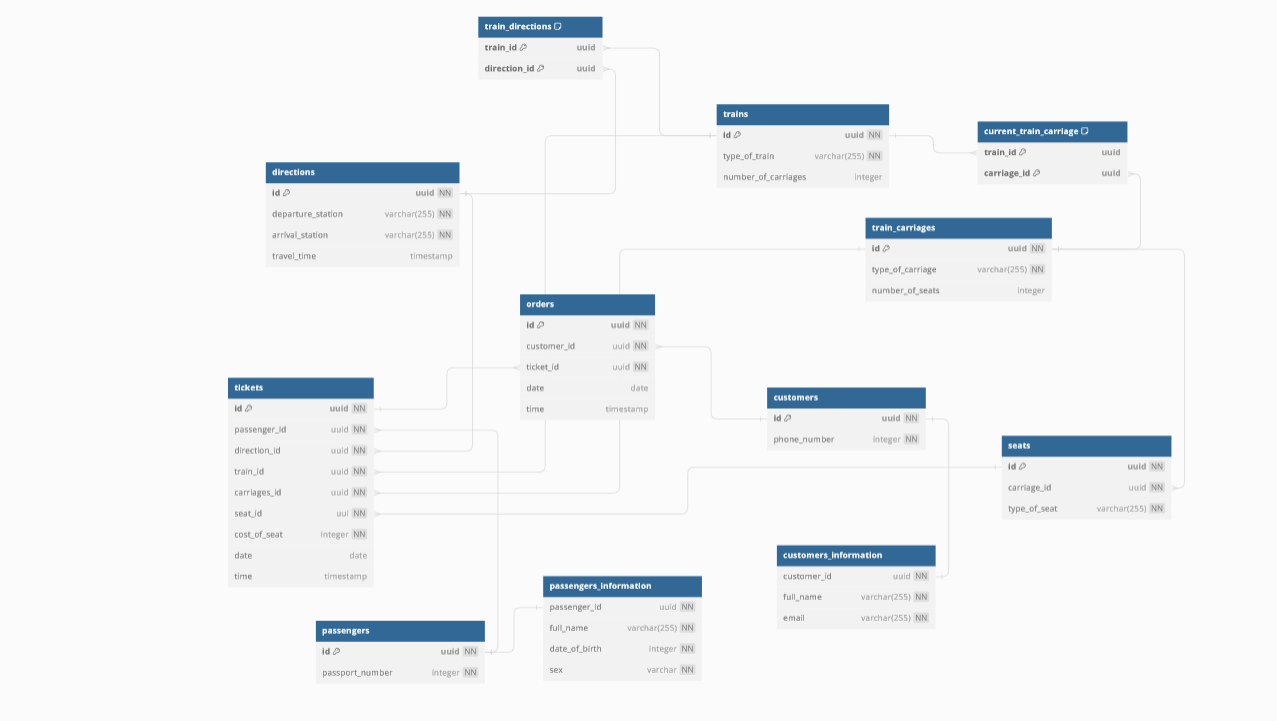
* Destruction
* Adding Noise
* Randomization
* Faking
* Advanced Faking
* Pseudonymization
* Generic Hashing
* Partial scrambling

1. Предоставить отчёт, включить в него следующие данные:

* перечень таблиц и полей, задействованных в Materialized Views;
* данные из Materialized Views;
* код по маскировке и анонимизации.

На защите лабораторной необходимо будет продемонстрировать, как работает маскировка на ваших данных, Materialized Views, сами данные из БД, а также код.

Моя, спроектированная в первой лабораторной работе, ER-diagram(а):

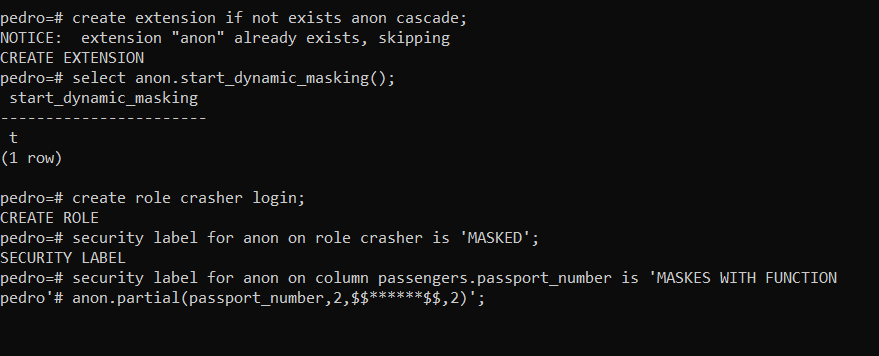


* 1. Динамическая маскировка данных.

В рамках своей модели, я решила маскировать и анонимизировать конфиденциальные данные или информацию, позволяющую установить личность (персональные данные), в моем случае это: ФИО, номера паспортов, номера телефонов, почтовые адреса.

Режим «MASKED» ставим для пользователя crasher.

* Маскировка данных таблицы passengers(passport\_number):





* Маскировка данных таблицы passengers\_information (full\_name)



* Маскировка данных таблицы customers(phone\_number):

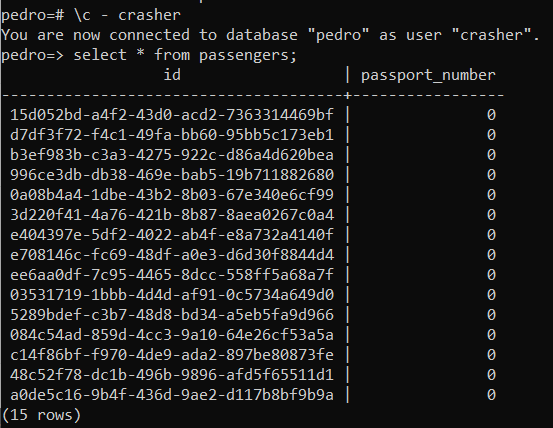


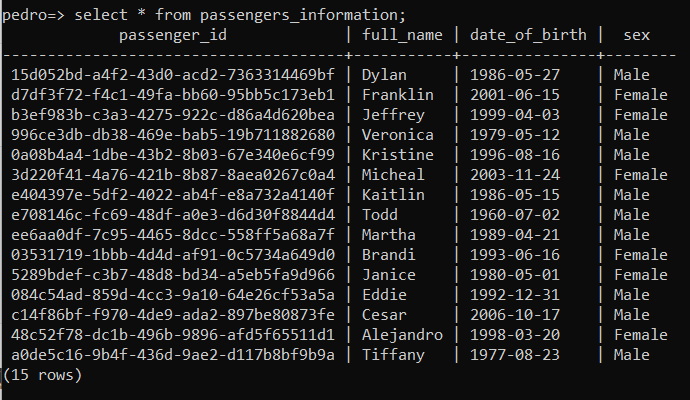
* Маскировка данных таблицы customers\_information(full\_name, email):

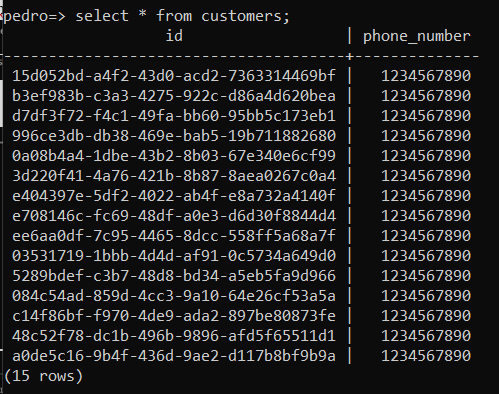


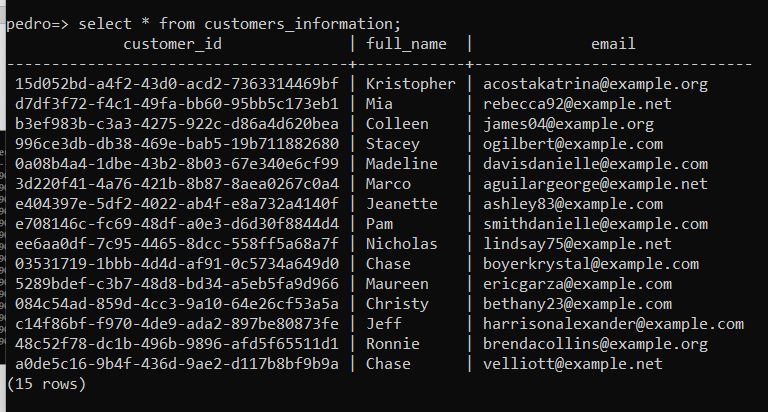


Проверка работоспособности (подключилась от лица пользователя crasher):



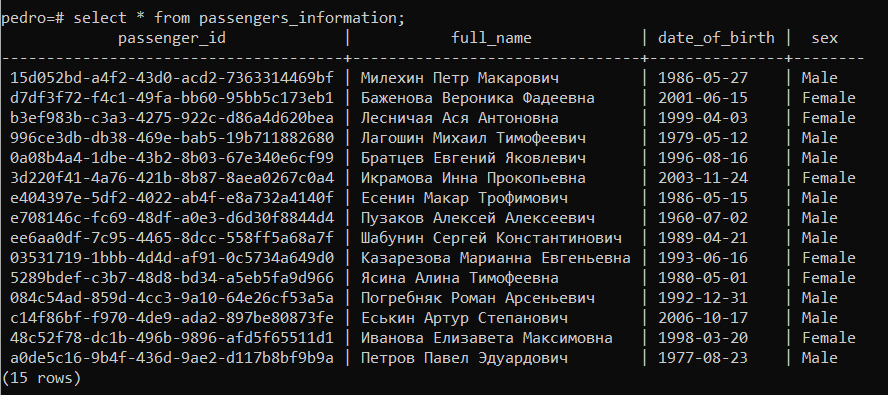




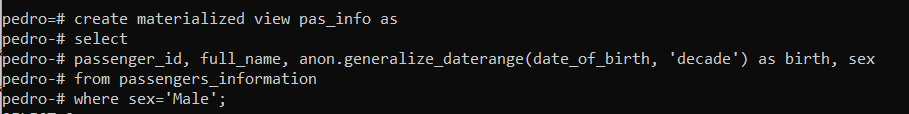


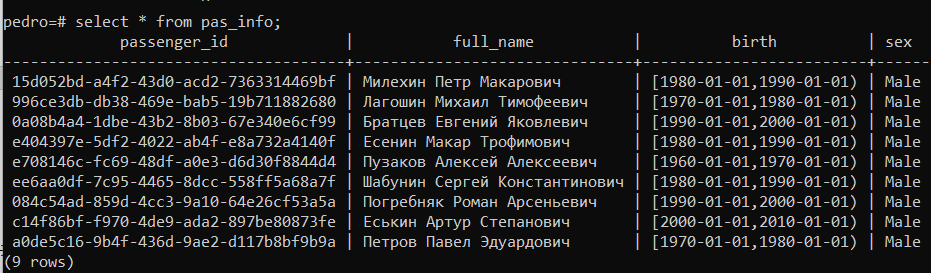
* 1. Создание Materialized Views.
* Использование Generalization:

Наиболее подходящей таблицей, на основе которой я могу создать материализованное представление, используя метод маскирования Generalization, является таблица passengers\_information. Я решила, что создам представление, которое выведет данные пассажиров только мужского пола, при этом дата их рождения будет «обобщена».



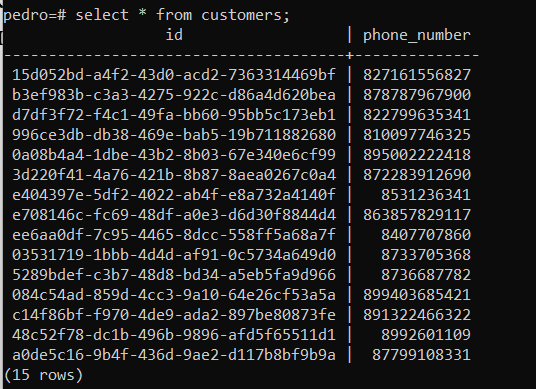
Код создания представления:

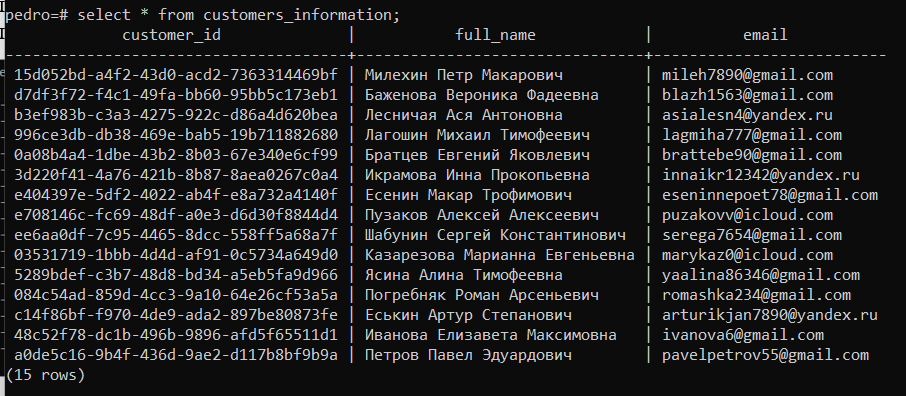


Результат вызова представления: 

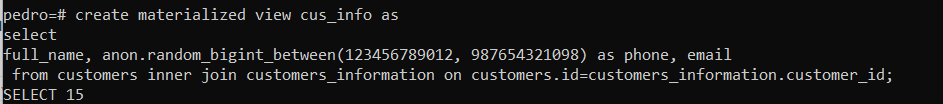
* Использование Randomization:

Я решила создать материализованное представление на основе таблиц customers и customers\_information, где у меня будут скрыты настоящие номера телефонов покупателей (за счет рандомизации номеров) и при этом выведена вся информация о покупателях (связь таблиц через customer\_id).

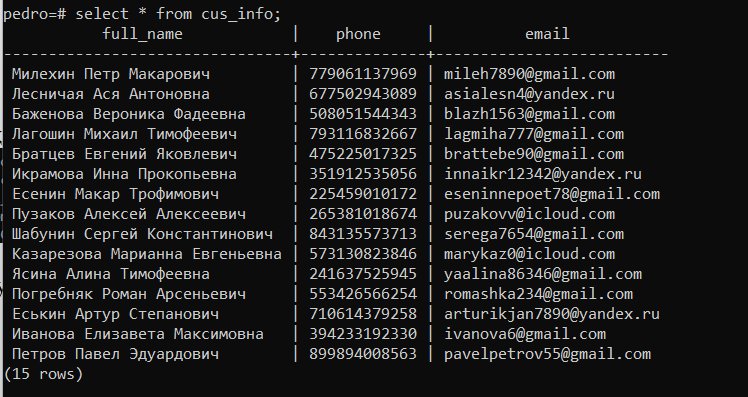




Код создания представления:

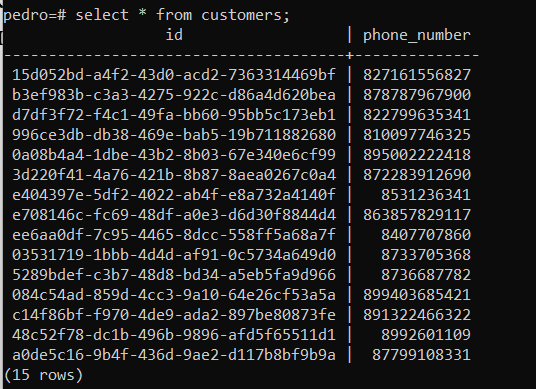


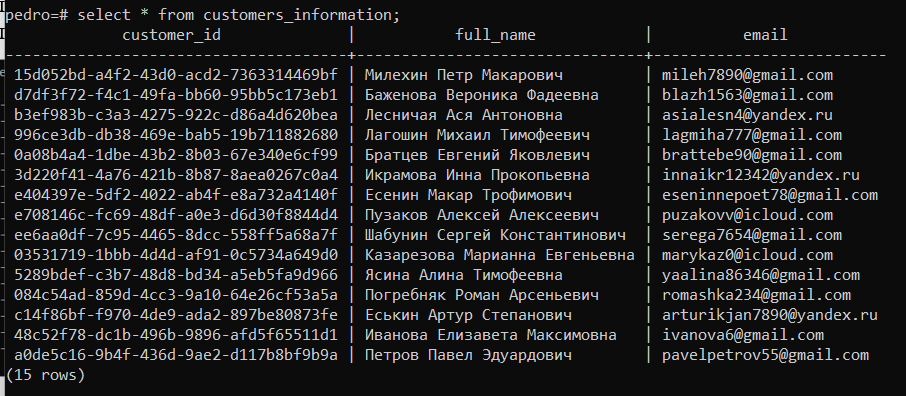
Результат вызова представления:

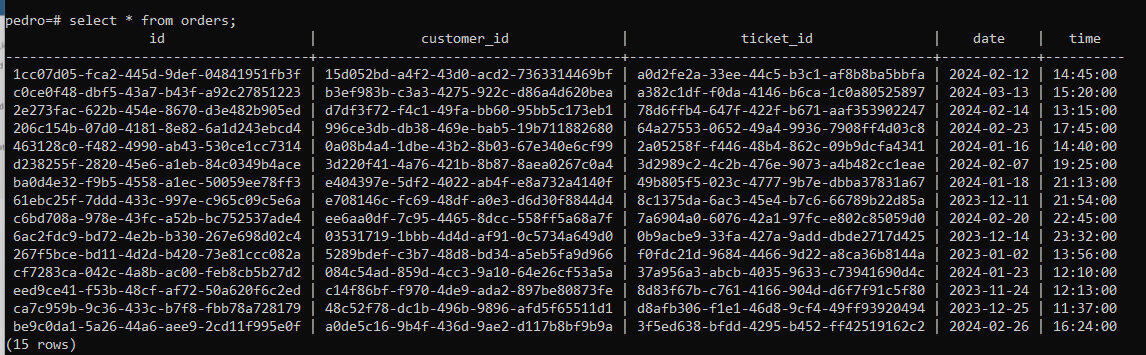


* Использование Partial Scrambling:

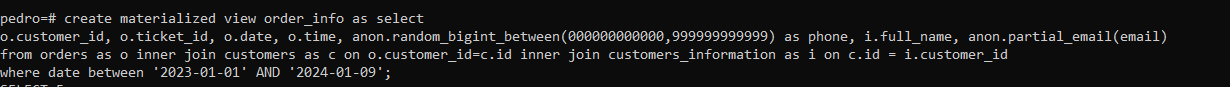
Я решила создать материализованное представление на основе таблиц customers, customers\_information и orders, где у меня будут частично скрыты почтовые адреса (ну и для красоты рандомизированы телефонов покупателей, как в предыдущем пункте) и выведена вся информация о заказе конкретных покупателей (связь таблиц через customer\_id), при этом будут выведены только заказы, сделанные в период с 01.01.23 по 09.01.24.







Код создания представления:



Результат вызова представления:

