

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ»
МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
(ФГБОУ ВО КубГМУ Минздрава России)

Задания (кейс-практика)

для завершения формирования практических навыков и оценки уровня
цифровых компетенций в рамках обучения на цикле дополнительного
профессионального образования (профессиональной переподготовки)
«Основы анализа больших данных в медицине» для присвоения
квалификации «Специалист по большим данным»

Краснодар, 2025

Кейс 1.

1. Создать структуру базы данных, состоящую из таблиц:

Таблица	Атрибут
Пациент	
	Фамилия
	Имя
	Отчество
	Пол
	Дата рождения
	Номер полиса ОМС
Врач	
	Фамилия
	Имя
	Отчество
	СНИЛС
	Специальность (использовать справочник специальностей)
Случаи оказания медицинской помощи	
	Дата начала
	Дата окончания
	Ответственный врач
	Результат (справочник: выздоровление, улучшение, динамическое наблюдение, направление в стационар, ухудшение, летальный исход)
Диагноз, установленный в рамках случая оказания медицинской помощи	
	Дата установления
	Код диагноза(по справочнику МКБ 10)
	Тип диагноза (основной, сопутствующий)
Медицинские услуги, оказанные в рамках случая	
	Дата оказания услуги
	Код услуги (справочник номенклатуры)
	Врач
Справочник диагнозов	

	Код МКБ
	Наименование
Справочник специальностей	
	Код специальности
	Наименование
Справочник номенклатуры услуг	
	Код услуги
	Наименование

Наименование таблиц и полей, типы данных установить в соответствии с принятыми правилами формирования структуры БД. Установить в каждой таблице первичный ключ.

2. Для таблиц-справочников загрузить содержимое из файлов средствами DBeaver.

В начале выполните скрипт create_table.sql, который создаст таблицы:

Справочник диагнозов с названием таблицы МКБ.

Справочник специальностей с названием таблицы rbSpeciality.

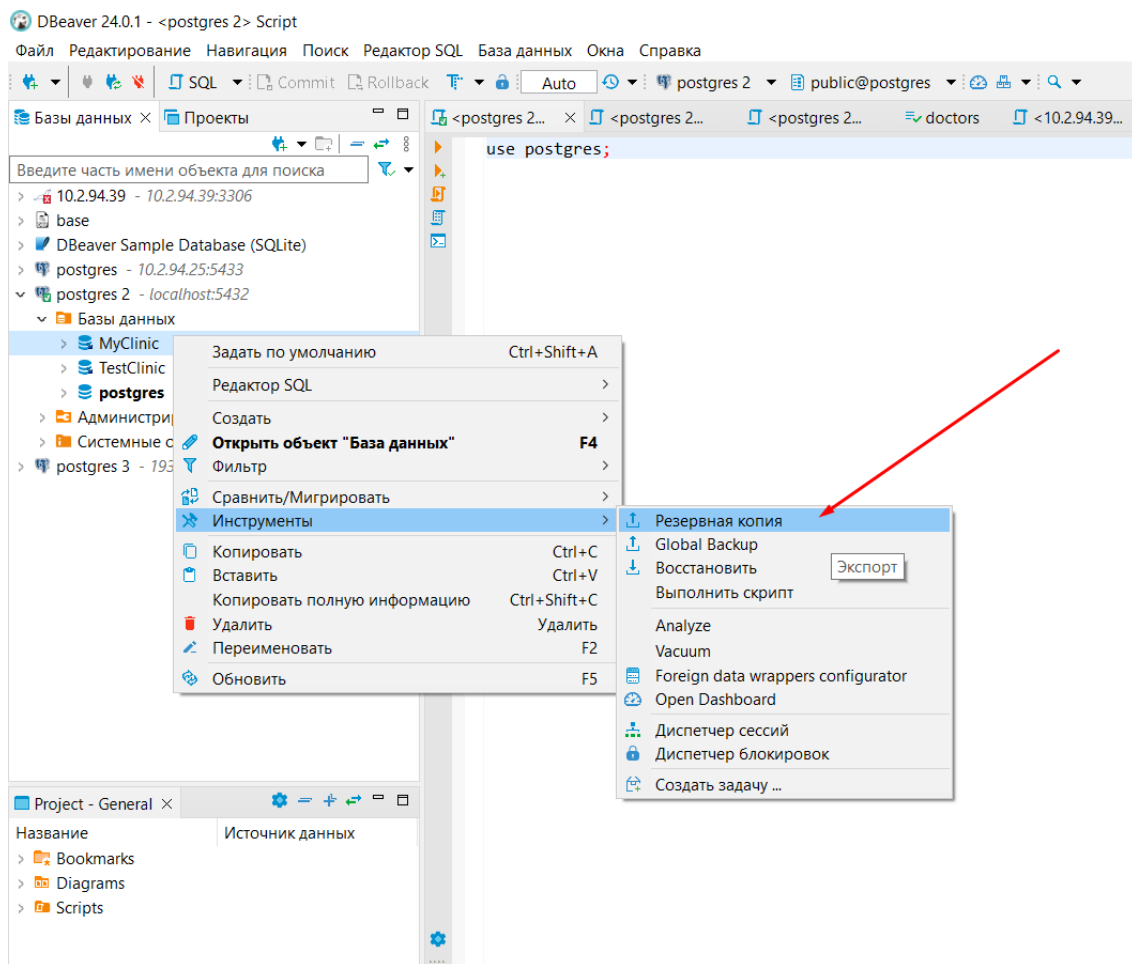
Справочник номенклатурных услуг с названием таблицы rbService.

Далее выполните следующие sql скрипты для заполнения справочников:

- Spec.sql
- МКБ.sql
- Usl.sql

Справочник	Файл
Справочник специальностей	Spec.sql
Справочник МКБ	МКБ.sql
Справочник номенклатуры услуг	Usl.sql

3. Результат разработки и дальнейшей обработки в Модуле 5 выгрузить в резервную копию:



Кейс 2.

Наполнение базы данных из Кейса 1 посредством файлов с данными.

1. Необходимо написать модуль, реализующий функционал по открытию файла Персона.csv и его и загрузке в соответствующую таблицу Пациенты базы данных, созданной в Кейсе 1.
2. Необходимо написать модуль, реализующий функционал по открытию файла Услуги.csv и его и загрузке в соответствующую таблицу «Случаи оказания медицинской помощи» базы данных, созданной в Кейсе 1.
3. Необходимо написать модуль, реализующий функционал по открытию файла Услуги.csv и его и загрузке в соответствующую таблицу «Диагнозы, установленный в рамках случая» базы данных, созданной в Кейсе 1.
4. Необходимо написать модуль, реализующий функционал по открытию файла Услуги.csv и его и загрузке в соответствующую таблицу «Медицинские услуги» базы данных, созданной в Кейсе 1.
5. Необходимо написать модуль для осуществления контроля загруженных из файлов данных. Данный модуль должен обеспечивать функционал для контроля количества загруженных строк в соответствии с количеством строк в файлах. Также модуль должен обеспечивать функционал по контролю корректности построенных связей между таблицами по ключевым полям.

6. Необходимо написать модуль, обеспечивающий посредством SQL запросов вывод данных в терминал из таблицы Пациенты, в том числе с использованием предикатов по возрасту и полу. Также в модуле должна быть обеспечена функциональность по группировке пациентов (исключая вывод персональных данных) по возрасту и полу с выводом результата в терминал. Группировка по возрасту должна быть обеспечена с шагом 10 лет.
7. Необходимо написать модуль, обеспечивающий посредством SQL запросов вывод данных в терминал из таблицы «Случаи оказания медицинской помощи», в том числе с использованием предикатов по периоду лечения и исходу обращения. Также в модуле должна быть обеспечена функциональность по группировке случаев по месяцу и году или исходу обращения с выводом результата в терминал.
8. Необходимо написать модуль, обеспечивающий посредством SQL запросов вывод данных в терминал из таблицы «Диагнозы, установленный в рамках случая», в том числе с использованием предикатов по коду МКБ. Также в модуле должна быть обеспечена функциональность по группировке случаев по коду МКБ с выводом результата в терминал.
9. Необходимо написать модуль, обеспечивающий посредством SQL запросов вывод данных в терминал из таблицы «Медицинские услуги», в том числе с использованием предикатов по коду услуги. Также в модуле должна быть обеспечена функциональность по группировке случаев по коду услуги с выводом результата в терминал.
10. Необходимо обеспечить возможность сохранения результатов выборок в формат CSV с заголовками.

Кейс 3.

Блок 1. Определение строк для выборки.

1. Для дальнейшей работы вам необходимо использовать два файла «услуги.csv», и «персона.csv», данные из этих файлов необходимо загрузить в табличный редактор.
2. Вам необходимо определить какие строки из файла «услуги.csv» должны использоваться для выполнения вашего задания по модулю 3, далее эти строки мы будем называть выборкой.
3. Для определения выборки вам необходимо взять первую букву вашей фамилии и далее оставить только те строки, которые указаны ниже в таблице. Данные строки мы будем называть **выборкой**.
Правило формирования выборки в таблице указано для первых 6-и строк, для последующих строк необходимо использовать сдвиг на 6, т.е. если в вашу выборку указано что необходимо использовать строку 1, то вам необходимо использовать строки 1,7,13,19,25,31,37 и т.д.

Строки	А	Б	В	Г	Д	Е	Ж	З	И	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш / Щ	Э	Ю	Я
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0
3	1	1	1	1	1	1	0	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1
4	1	1	1	0	0	0	1	1	1	0	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1
5	1	0	0	1	1	0	1	1	0	1	1	1	0	1	1	1	1	0	0	1	1	0	1	1	0	1	1
6	0	1	0	1	0	1	1	0	1	1	1	0	1	1	1	1	0	1	0	1	0	1	1	0	1	1	1

Пример: Допустим ваша фамилия начинается на букву «Ш», в этом случае вам необходимо использовать строки 2,4,5 и далее $8=2+6$, $10=4+6$, $11=5+6$. Суммарно у вас должно остаться 16987 строк.

(!) Важное: Выборка распространяется только на Кейс 3, для остальных кейсов необходимо использовать все данные из файлов.

Блок 2. Проверка данных

Вам необходимо найти и исключить из вашей Выборки строки услуг с ошибками. Под ошибкой подразумевается – некорректно указана дата рождения. Все дальнейшие расчеты в модуле 3 необходимо проводить только с оставшимися строками.

Пример: Для фамилии, которая начинается на букву «Ш» количество строк с ошибками равно 453. Значит в нашей выборке должно остаться $16534 \text{ строки} = 16987 - 453$. В

Блок 3. Пять вопросов

1) Попали ли в вашу выборку услуги с диагнозами

- а. А09.0
- б. В35.4
- с. С20

Ответ: 3 значения ДА/НЕТ

2) Какое количество услуг попало в вашу выборку по следующим разрезам:

- а. Мужчины, до 30-и лет
- б. Мужчина, от 30-и до 50-и лет
- с. Мужчина от 50-и лет.
- д. Женщины, до 30-и лет
- е. Женщины, от 30-и до 50-и лет
- ф. Женщины от 50-и лет

Ответ: 6 чисел

3) Какое количество персон (уникальных пациентов), попали в вашу выборку?

Ответ: Число

4) Дата последнего приема любого пациента, соответствующему критериями:

- а. Женщина старше 50-ти лет.

б. Мужчина до 30-и лет

с. Пациент (пол не важен) от 30 до 50-и лет.

Ответ: 3 числа

Кейс 4.

1. Определить категории ПДн, содержащиеся в выборке
2. Считать, что количество субъектов ПДн в общей БД больше, чем в выборке в $\times 1\,000$, $\times 10\,000$, $\times 50\,000$ (три варианта)
3. Определить уровень защищённости ПДн общей БД
4. В соответствии с Приказом ФСТЭК России от 18 февраля 2013 г. N 21 определить базовый набор мер для соответствующего уровня защищенности ПДн.