

Business Intelligence Project

Авторы:

Шейко Екатерина - <https://www.linkedin.com/in/katherine-sheiko>

Лебедева Наталия - <https://www.linkedin.com/in/>

Якута Вероника - <https://www.linkedin.com/in/veronika-yakuta-7020921b4>

Содержание проекта

- ◆ *Концептуальное проектирование БД*
- ◆ *Логическое проектирование БД*
- ◆ *Физическое проектирование БД*
- ◆ *Заполнение тестовыми данными*
- ◆ *Результаты тестирования БД (баг-репорт)*
- ◆ *ETL пакет*
- ◆ *Визуализация данных и итоги*

Данные проекта

"Personal Key Indicators of Heart Disease"

(Персональные ключевые показатели сердечно-сосудистых заболеваний)



Personal Key Indicators of Heart Disease

2020 annual CDC survey data of 400k adults related to their health status

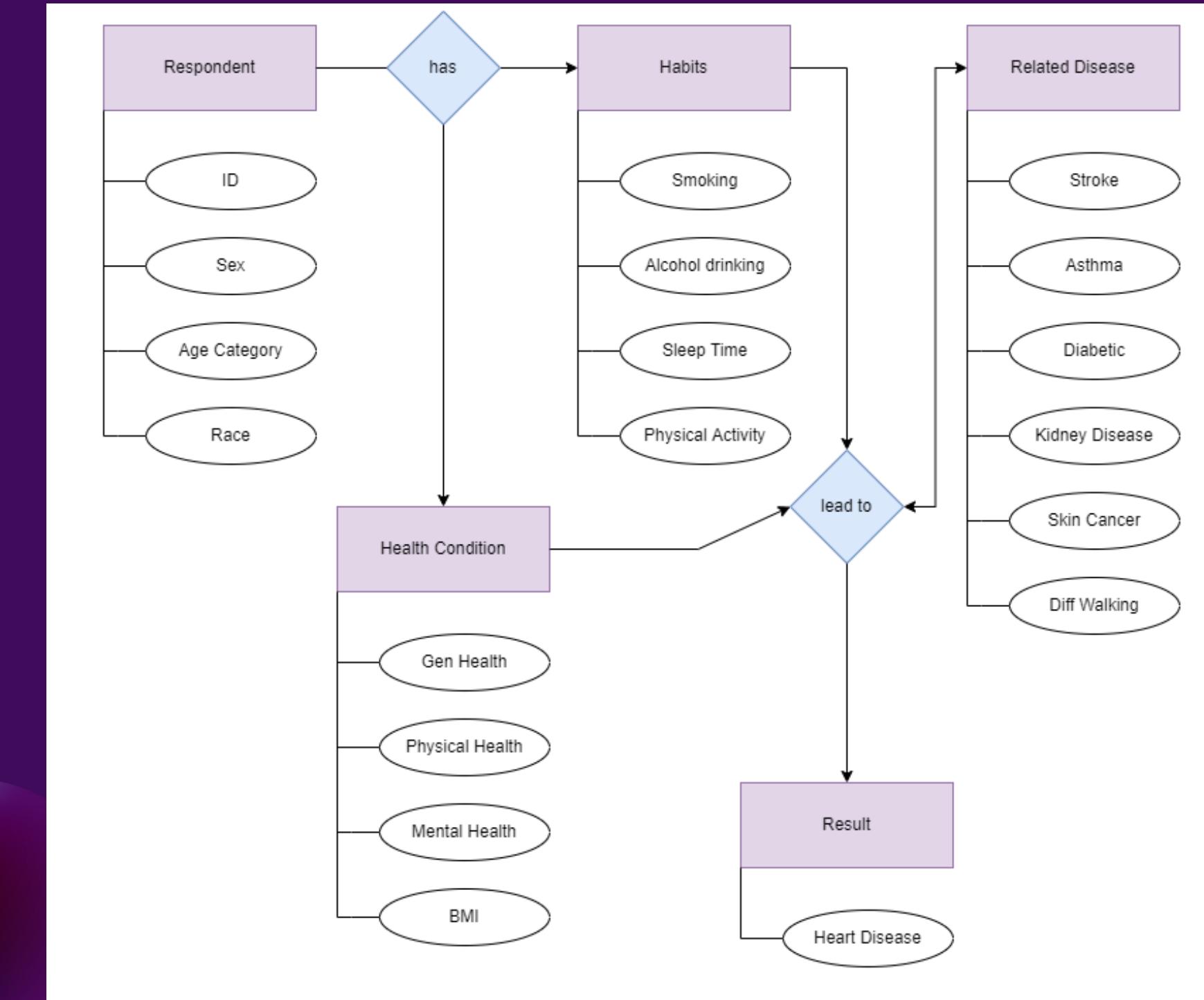
[kaggle.com](https://www.kaggle.com)



Для перехода к данным нажмите на картинку

01

Концептуальное проектирование БД



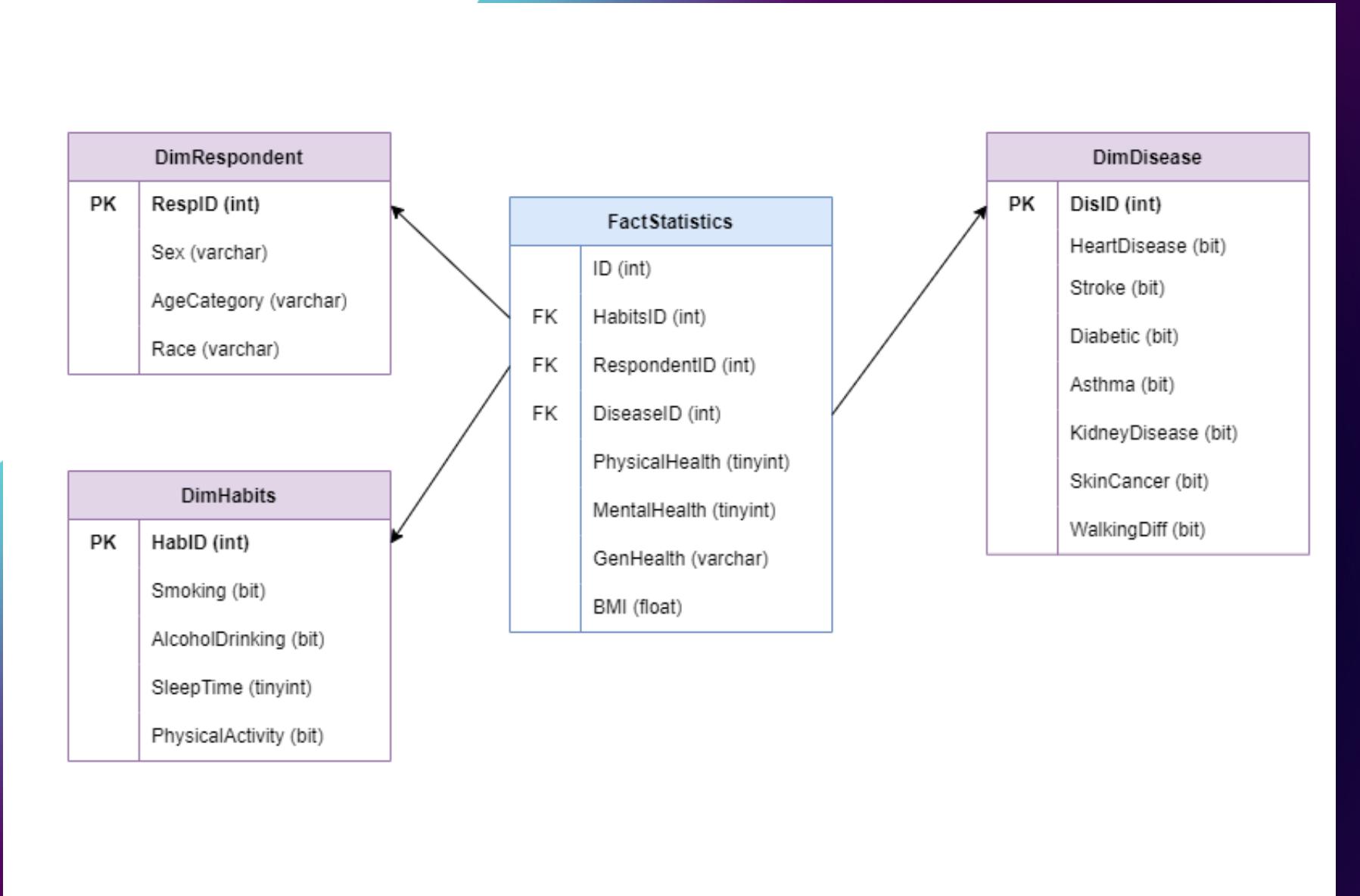


Схема "Звезда"

02 Логическое проектирование Бд

03

Физическое проектирование

Создание таблиц

```
CREATE TABLE [dbo].[FactStatistics]
(
    [ID] [INT] IDENTITY (1, 1) NOT NULL PRIMARY KEY,
    [HabitsID] [INT] NOT NULL,
    [RespondentID] [INT] NOT NULL,
    [DiseaseID] [INT] NOT NULL,
    [PhysicalHealth] [TINYINT] NOT NULL,
    [MentalHealth] [TINYINT] NOT NULL,
    [GenHealth] [VARCHAR] (100) NOT NULL,
    [BMI] [FLOAT] NOT NULL
) ON [PRIMARY]

CREATE TABLE [dbo].[DimRespondent]
(
    [RespID] [INT] IDENTITY (1, 1) NOT NULL PRIMARY KEY,
    [Sex] [VARCHAR] (100) NOT NULL,
    [AgeCategory] [VARCHAR] (100) NOT NULL,
    [Race] [VARCHAR] (100) NOT NULL
) ON [PRIMARY]

CREATE TABLE [dbo].[DimHabits]
(
    [HabID] [INT] IDENTITY (1, 1) NOT NULL PRIMARY KEY,
    [Smoking] [BIT] NOT NULL,
    [AlcoholDrinking] [BIT] NOT NULL,
    [SleepTime] [TINYINT] NOT NULL,
    [PhysicalActivity] [BIT] NOT NULL,
) ON [PRIMARY]

CREATE TABLE [dbo].[DimDisease]
(
    [DisID] [INT] IDENTITY (1, 1) NOT NULL PRIMARY KEY,
    [HeartDisease] [BIT] NOT NULL,
    [Stroke] [BIT] NOT NULL,
    [Diabetic] [BIT] NOT NULL,
    [Asthma] [BIT] NOT NULL,
    [KidneyDisease] [BIT] NOT NULL,
    [SkinCancer] [BIT] NOT NULL,
    [WalkingDiff] [BIT] NOT NULL,
) ON [PRIMARY]
```

Физическое проектирование

Добавление ограничений

```
ALTER TABLE [dbo].[FactStatistics]
ADD CONSTRAINT FK_HabitsID
FOREIGN KEY ([HabitsID]) REFERENCES [dbo].[DimHabits]([HabID]) ON DELETE CASCADE ON UPDATE CASCADE;

ALTER TABLE [dbo].[FactStatistics]
ADD CONSTRAINT FK_RespondentID
FOREIGN KEY ([RespondentID]) REFERENCES [dbo].[DimRespondent]([RespID]) ON DELETE CASCADE ON UPDATE CASCADE;

ALTER TABLE [dbo].[FactStatistics]
ADD CONSTRAINT FK_DiseaseID
FOREIGN KEY ([DiseaseID]) REFERENCES [dbo].[DimDisease]([DisID]) ON DELETE CASCADE ON UPDATE CASCADE;

ALTER TABLE [dbo].[FactStatistics]
ADD CONSTRAINT CK_PhysicalHealth
CHECK ([PhysicalHealth] >= 0 AND [PhysicalHealth] <= 30);

ALTER TABLE [dbo].[FactStatistics]
ADD CONSTRAINT CK_MentalHealth
CHECK ([MentalHealth] >= 0 AND [MentalHealth] <= 30);

ALTER TABLE [dbo].[DimRespondent]
ADD CONSTRAINT CK_Sex
CHECK ([Sex] = 'Male' OR [Sex] = 'Female');

ALTER TABLE [dbo].[FactStatistics]
ADD CONSTRAINT CK_BMI
CHECK ([BMI] > 0);

ALTER TABLE [dbo].[DimHabits]
ADD CONSTRAINT CK_SleepTime
CHECK ([SleepTime] >= 0 AND [SleepTime] <= 24);
```

04

Заполнение таблиц тестовыми данными

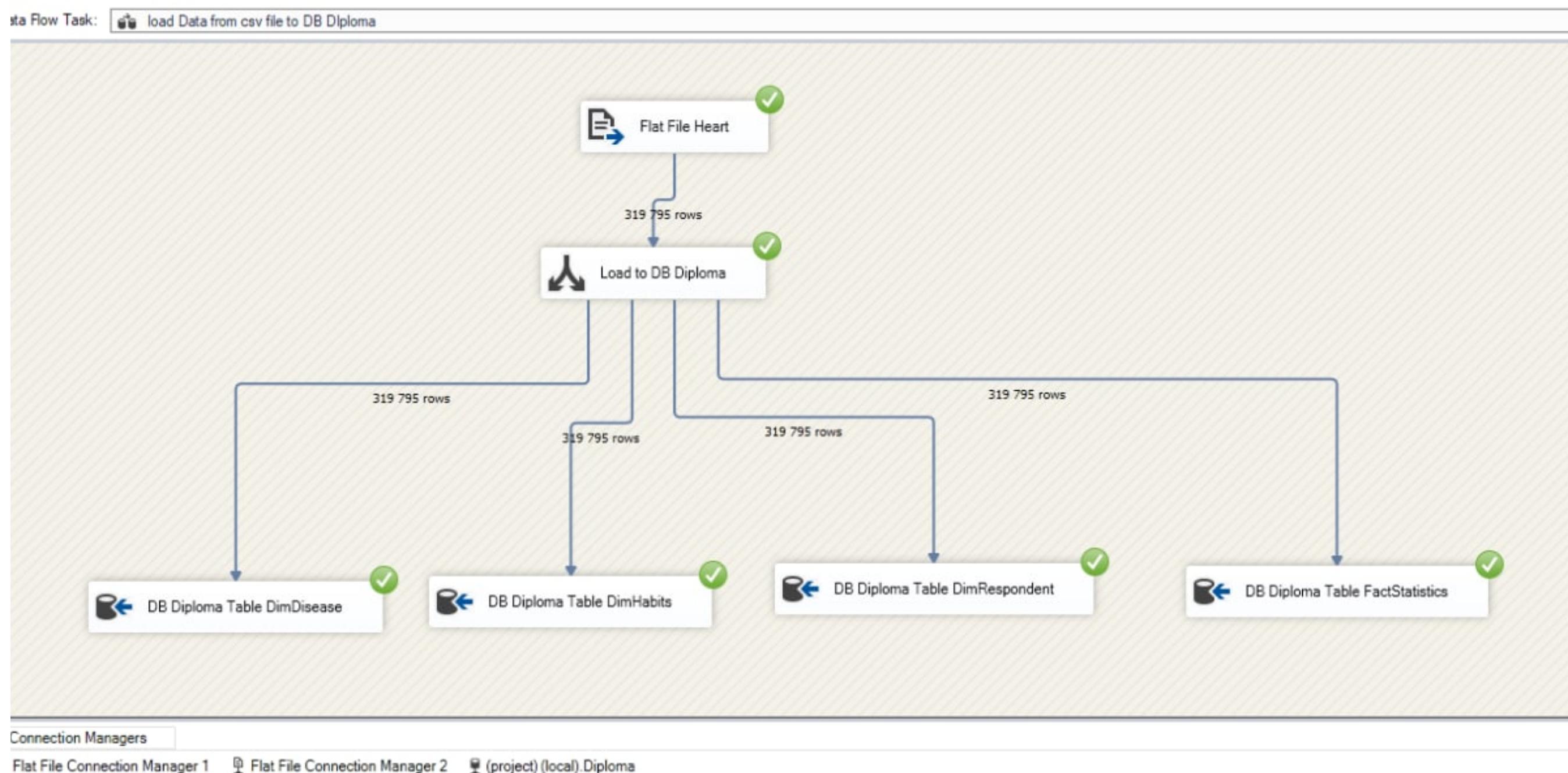
Таблица FactStatistics

```
declare @i int =1
while @i < 200 begin
INSERT INTO [dbo].[FactStatistics]
(
[HabitsID],
[RespondentID],
[DiseaseID],
[PhysicalHealth],
[MentalHealth],
[GenHealth],
[BMI]
)
VALUES
((SELECT TOP 1 [HabID] FROM [DimHabits] ORDER BY NEWID())
,(SELECT TOP 1 [RespID] FROM [DimRespondent] ORDER BY NEWID())
,(SELECT TOP 1 [DisID] FROM [DimDisease] ORDER BY NEWID())
,(SELECT RAND()*(30-0)+0)
,(SELECT RAND()*(30-0)+0)
,"Good"
,(SELECT RAND()*(100-1)+1))
set @i = @i + 1
end
```

05 Результаты тестирования (баг-репорт)

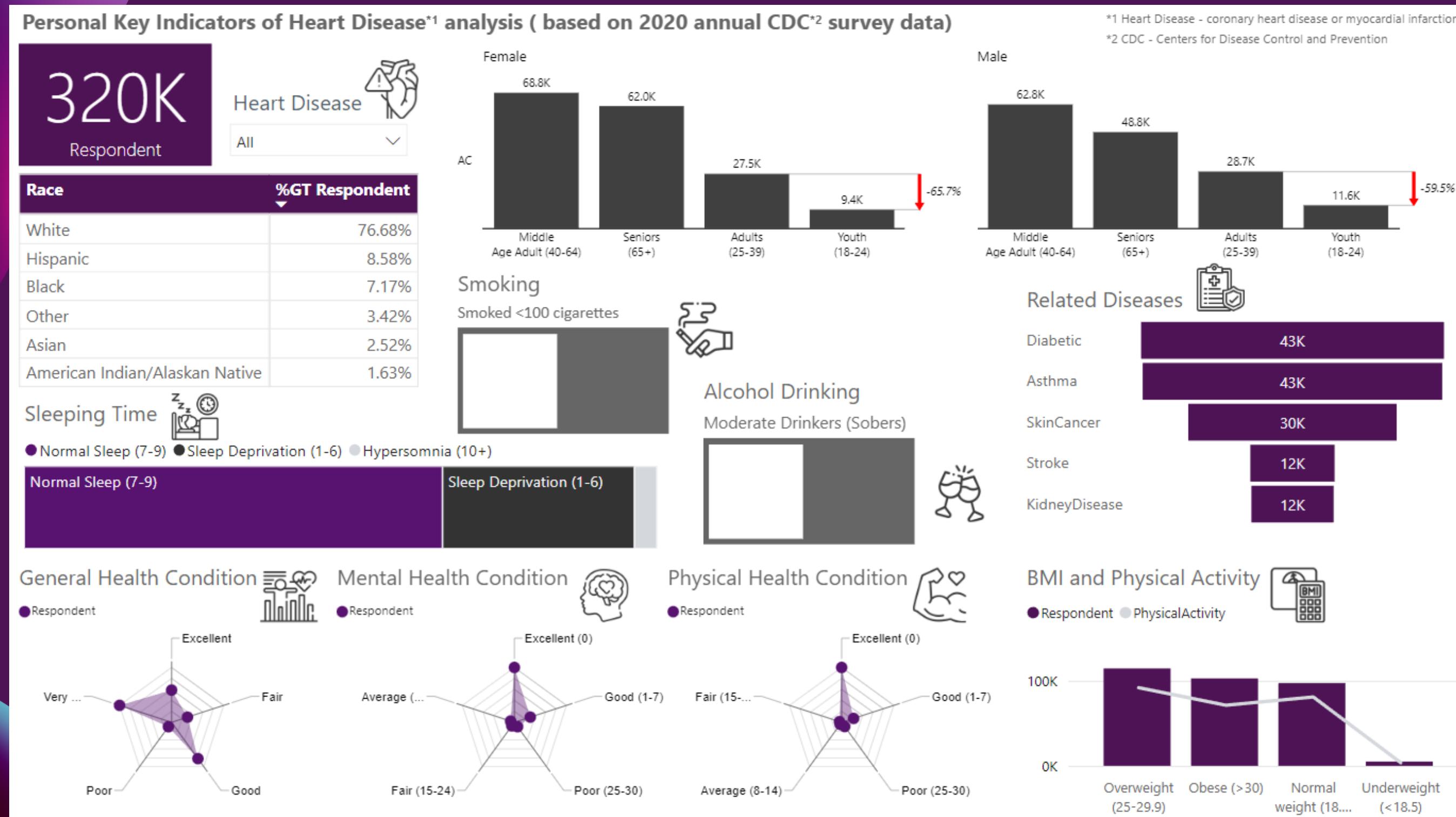
--ожидается добавление записи <code>INSERT INTO [dbo].[DimRespondent] VALUES ('Female', 35, 'White')</code> --ожидается срабатывание проверки на пол <code>INSERT INTO [dbo].[DimRespondent] VALUES ('M', 45, 'White')</code>	(1 row affected) Msg 547, Level 16, State 0, Line 3 Конфликт инструкции INSERT с ограничением CHECK "CK_Sex". Конфликт произошел в базе данных "Diploma", таблица "dbo.DimRespondent", column 'Sex'. Выполнение данной инструкции было прервано.
--ожидается срабатывание проверки на допустимые значения от 0 до 30 <code>INSERT INTO [dbo].[FactStatistics] ([PhysicalHealth]) VALUES (80)</code>	Msg 547, Level 16, State 0, Line 10 Конфликт инструкции INSERT с ограничением CHECK "CK_PhysicalHealth". Конфликт произошел в базе данных "Diploma", таблица "dbo.FactStatistics", column 'PhysicalHealth'. Выполнение данной инструкции было прервано.
--ожидается срабатывание проверки на допустимые значения от 0 до 30 <code>INSERT INTO [dbo].[FactStatistics] ([MentalHealth]) VALUES (60)</code>	Msg 547, Level 16, State 0, Line 13 Конфликт инструкции INSERT с ограничением CHECK "CK_MentalHealth". Конфликт произошел в базе данных "Diploma", таблица "dbo.FactStatistics", column 'MentalHealth'. Выполнение данной инструкции было прервано.
--ожидается срабатывание проверки на допустимые значения > 0 <code>INSERT INTO [dbo].[FactStatistics] ([BMI]) VALUES (-10)</code>	Msg 547, Level 16, State 0, Line 16 Конфликт инструкции INSERT с ограничением CHECK "CK_BMI". Конфликт произошел в базе данных "Diploma", таблица "dbo.FactStatistics", column 'BMI'. Выполнение данной инструкции было прервано.
--ожидается срабатывание проверки на допустимые значения от 0 до 24 <code>INSERT INTO [dbo].[DimHabits] ([SleepTime]) VALUES (25)</code>	Msg 547, Level 16, State 0, Line 26 Конфликт инструкции INSERT с ограничением CHECK "CK_SleepTime". Конфликт произошел в базе данных "Diploma", таблица "dbo.DimHabits", column 'SleepTime'. Выполнение данной инструкции было прервано.

Создание ETL пакета



06

С использованием Visual Studio (SSIS)



Визуализация данных

Спасибо за внимание!