|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра практической и прикладной информатики (ППИ)**

**КУРСОВАЯ РАБОТА**

по дисциплине «Технологии организации обработки и хранения статистических данных»

**Тема курсовой работы**: «Анализ признаков и рисков сахарного диабета для раннего прогнозирования»

|  |  |  |
| --- | --- | --- |
| Студент группы ИМБО-02-22 | Ким Кирилл Сергеевич | (подпись) |
| Руководитель  курсовой работы | доцент, к.п.н. Митина О.А. | (подпись) |
| Работа представлена к защите | «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_2023 г. |  |
| Допущен к защите | «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_2023 г. |  |

Москва 2023 г.

|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

|  |  |  |
| --- | --- | --- |
|  |  | Утверждаю |
|  |  | Заведующий кафедрой ПМ |
|  |  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Смоленцева Т.Е.  (подпись) |
|  |  | «21» сентября 2023 г. |

**ЗАДАНИЕ**

**на выполнение курсовой работы**

по дисциплине «Технологии организации обработки и хранения статистических данных»

|  |  |  |
| --- | --- | --- |
| Студент Ким Кирилл Сергеевич | | Группа ИМБО-02-22 |
| **Тема** «Анализ признаков и рисков сахарного диабета для раннего прогнозирования» | | |
| **Исходные данные:** выбранный студентом данные | | |
| **Перечень вопросов, подлежащих разработке, и обязательного графического материала:**  Характеристика деятельности организации и общее описание признаков, влияющих на развитие сахарного диабета (описание признаков, типов данных, общее текстовое описание сахарного диабета и графическое влияния признаков на риск)  Графические модели зависимости влияния признаков на риск сахарного диабета (описание путем представления графиков по признакам)  Анализ риска возникновения сахарного диабета (выявление зависимостей между признаками)  Раннее прогнозирование сахарного диабета | | |
| **Срок представления к защите курсовой работы: до** «15» декабря 2023 г. | | |
| **Задание на курсовую работу выдал** | *Подпись руководителя* | Митина О.А.  *(ФИО руководителя)* |
| «21» сентября 2023 г. | | |
| **Задание на курсовую работу получил** | *Подпись обучающегося* | Ким К.С.  *(ФИО обучающегося)* |
| «21» сентября 2023 г. | | |

**ОТЗЫВ**

**на курсовую работу**

**по дисциплине «Технологии организации, обработки и хранения статистических данных»**

**Студент Ким Кирилл Сергеевич** **ИМБО-02-22**

(ФИО студента) (Группа)

Характеристика курсовой работы

|  |  |  |  |
| --- | --- | --- | --- |
| Критерий | Да | Нет | Не полностью |
| 1. Соответствие содержания курсовой работы указанной теме |  |  |  |
| 2. Соответствие курсовой работы заданию |  |  |  |
| 3. Соответствие рекомендациям по оформлению текста, таблиц, рисунков и пр. |  |  |  |
| 4. Полнота выполнения всех пунктов задания |  |  |  |
| 5. Логичность и системность содержания курсовой работы |  |  |  |
| 6. Отсутствие фактических грубых ошибок |  |  |  |

Замечаний:

Рекомендуемая оценка:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_

Подпись руководителя ФИО руководителя

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 1](#_Toc148772566)

[1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ 2](#_Toc148772567)

[1.1 Сахарный диабет 2](#_Toc148772568)

[1.2 Корреляционно-регрессионный анализ 9](#_Toc148772569)

[2 ПРАКТИЧЕСКАЯ ЧАСТЬ 14](#_Toc148772570)

[2.1 Корреляционно-регрессионный анализ признаков и рисков сахарного диабета для раннего прогнозирования. 14](#_Toc148772571)

[ЗАКЛЮЧЕНИЕ 26](#_Toc148772572)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 27](#_Toc148772573)

[Теоретическая часть 27](#_Toc148772574)

[Практическая часть 28](#_Toc148772575)

[ПРИЛОЖЕНИЯ 29](#_Toc148772576)

[Приложение А 30](#_Toc148772577)

# ВВЕДЕНИЕ

В современном мире люди все чаще подвержены риску сахарного диабета, что требует всестороннего изучения связей между биологическими признаками человека.

Таким образом, тема данной курсовой работы является актуальной для медицины и человека в целом.

Целью данной курсовой работы является – провести анализ и построить модель прогнозирования риска сахарного диабета.

Задачи, решаемые в данной курсовой работе:

1. Изучение научной и методической литературы о сахарном диабета;

2. Поиск/Сбор данных о сахарном диабете;

3. При необходимости подготовить данные, провести очистку и предобработку;

4. Проанализировать исходные данные, построив графики зависимостей;

5. Используя знания математической статистики и современных средств обработки данных.

# 1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

## Сахарный диабет

Сахарный диабет – это хроническое заболевание эндокринной системы человека, характеризующиеся длительным повышением концентрации глюкозы в крови и сопутствующими изменениями процессов обмена веществ. Представляет серьёзную угрозу для здоровья и жизни больного, так как провоцирует развитие тяжелых сопутствующих заболеваний. В основе сахарного диабета лежит инсулиновая недостаточность, что приводит к увеличению сахара в крови и появлению его в моче. Это происходит из –за того, что поджелудочная железа вырабатывает недостаточное количество гормона – инсулина, который регулирует углеводный обмен в организме. При наследственной предрасположенности к заболеванию его возникновение могут провоцировать переедание, злоупотребление сладким, нервное перенапряжение, вирусная инфекция. Диабет может развиваться после краснухи, кори, гриппа и других вирусных заболеваниях. Самостоятельно выявить заболевание невозможно, но некоторые отклонения в организме могут быть первыми тревожными сигналами. К таковым признакам можно отнести: частое и обильное мочеиспускание, ощущение постоянной жажды, кожный зуд, постоянные скачки весовых соотношений (масса тела то резко увеличивается, то моментально снижается), проявление кожных высыпаний. На начальной стадии болезни сахарный диабет практически незаметен, все симптомы усиливаются в тот момент, когда недуг активизировался. Диагностировать болезнь при первичном осмотре и беседе с пациентом невозможно, для того чтобы поставить точный диагноз нужно обязательно провести лабораторные исследования. К сожалению, несвоевременное

лечение сахарного диабета приводит к серьезным осложнениям для больного. Различают 2 типа сахарного диабета:

Сахарный диабет 1 типа (инсулин зависимый диабет) – это деструкция клеток поджелудочной железы, которая приводит к полной инсулиновой недостаточности. При 1 типе диабета практически все клетки поджелудочной железы, которые выделяют инсулин, разрушаются вследствие чего железа не в состоянии продуцировать инсулин. Инфекция сама по себе не разрушает клетки поджелудочной железы, она включает иммунную систему, клетки которой и уничтожают клетки поджелудочной. Данный тип составляет 10% от общего числа больных сахарным диабетом. Болеют дети и взрослые до 30 лет. При диабете 1 типа пациент вынужден постоянно вводить инсулин, который нужен для нормального передвижения глюкозы в организме.

Основной метод лечения – инъекции инсулина. Поэтому и называется 1 тип диабета – инсулин зависимый диабет. Этот тип диабета активно прогрессирует, быстро развиваются осложнения и стадия декомпенсации.

Сахарный диабет 2 типа (инсулин независимый) – это нарушение углеводного обмена с развитием гипергликемии. При данном типе поджелудочная железа не разрушается и продолжает вырабатывать инсулин, но в организме развивается резистентность (снижение чувствительности) клеток к инсулину. В результате этого в клетки не поступает нужного количества глюкозы, даже в присутствии инсулина. При наличии данного типа диабета у кого-то в роду, вероятность развития у потомка на протяжении жизни составляет 40%. Также, часто его развитию способствуют ожирение, перенапряжение и стрессовые ситуации. Его коррекция может быть достигнута диетой, снижением массы тела и сахароснижающими таблетками.

При диабете 2 типа не нужно вводить инсулин, так как проблема не в выработке инсулина, а именно в усвоении глюкозы тканями. Но, по мере прогрессирования диабета, выделение инсулина клетками поджелудочной железы снижается и тогда приходиться назначать инсулин.

Сахарный диабет является сложным заболеванием, которое трудно поддается лечению. При его развитии в организме происходит нарушение углеводного обмена и снижение синтеза инсулина поджелудочной железой, в результате чего глюкоза перестает усваиваться клетками и оседает в крови в виде микрокристаллических элементов. Точные причины, по которым начинает развиваться данный недуг, ученым установить до сих пор не удалось. Но благодаря им были выявлены факторы риска, которые могут спровоцировать возникновение этого заболевания у людей разностной возрастной категории.

Прежде чем рассматривать факторы риска развития сахарного диабета, необходимо сказать, что это заболевание, как уже было сказано выше, имеет два типа, и каждый из них имеет свои особенности. Диабет 1 типа характеризуется системными изменениями в организме, при которых нарушается не только углеводный обмен, но и функциональность поджелудочной железы. По каким – то причинам ее клетки перестают вырабатывать инсулин в нужном количестве, в результате чего сахар, проникающий в организм вместе с пищей, не подвергается процессам расщепления и, соответственно, не может усваиваться клетками.

Самыми распространенными осложнениями этой болезни являются следующие состояния:

* гипергликемия – повышение уровня сахара в крови за пределы нормы (свыше 7 ммоль/л);
* гипогликемия – снижение уровня глюкозы в крови за пределы нормы (ниже 3,3 ммоль/л);
* гипергликемическая кома – повышение уровня сахара в крови свыше 30 ммоль/л;
* гипогликемическая кома – снижение уровня глюкозы в крови ниже 2,1 ммоль/л;
* диабетическая стопа – снижение чувствительности нижних конечностей и их деформация;
* диабетическая ретинопатия – снижение остроты зрения;
* тромбофлебит – образование в стенках сосудов бляшек;
* гипертония – повышение артериального давления;
* гангрена – некроз тканей нижних конечностей с последующим развитием абсцесса;
* инсульт и инфаркт миокарда.

Основными факторами его развития являются:

* наследственная предрасположенность;
* вирусные заболевания;
* интоксикация организма;
* неправильное питание;
* частые стрессы.

В возникновении СД1 главную роль играет наследственная предрасположенность. Если кто-то из членов семьи страдает от этого недуга, то риски его развития у последующего поколения составляют примерно 10- 20%.

При этом следует отметить, что в данном случае речь идет не об установленном факте, а о предрасположенности. То есть если мать или отец болеют СД1, это вовсе не значит, что у их детей также будет диагностировано это заболевание. Предрасположенность говорит о том, что если человек не будет проводить профилактические мероприятия и будет вести неправильный образом жизни, то у него есть большая вероятность стать диабетиком в течение нескольких лет

Однако и в этом случае необходимо учитывать, что если от диабета страдают сразу оба родителя, то вероятность возникновения его у их ребенка значительно повышается. И нередко именно в таких ситуациях это заболевание диагностируется у детей еще в школьном возрасте, хотя они еще

не имеют вредных привычек и ведут активный образ жизни.

Вирусные заболевания – еще одна причина, по которой может развиваться СД1. Особенно опасными в этом случае являются такие болезни, как паротит и краснуха. Учеными уже давно было доказано, что эти заболевания негативно сказываются на работе поджелудочной железы и приводят к повреждению ее клеток, снижая, таким образом, уровень инсулина в крови.

Следует отметить, что в зоне риска не только уже рожденные дети, но и те, которые еще находятся в утробе матери. Любые вирусные заболевания, которые переносит беременная женщина, могут спровоцировать развитие у ее ребенка сахарного диабета 1 типа.

Многие люди, работающие на заводах и предприятиях, где используются химические вещества, также подвержены большому риску. Действие химикатов негативно сказывается на работе всего организма, в том числе и на функциональности поджелудочной железы.

Химиотерапии, которые проводятся для лечения различных онкологических заболеваний, также оказывают токсичное действие на клетки организма, поэтому их проведение тоже в несколько раз увеличивает вероятность развития СД1 у человека.

Неправильное питание является одним из самых распространенных причин развития сахарного диабета 1 типа. Ежедневный рацион современного человека содержит в себе огромное количество жиров и углеводов, что оказывает сильную нагрузку на пищеварительную систему, в том числе и на поджелудочную. Со временем ее клетки повреждаются, и синтез инсулина нарушается. Необходимо также отметить, что из-за неправильно питания СД1 может развиваться и у детей в возрасте 1-2 лет.

Симптомы обоих типов сахарного диабета проявляются тогда, когда уровень глюкозы в крови превышает 6,0 ммоль/л и тогда, когда глюкоза попадает в мочу, где ее в норме быть не должно.

Для сахарного диабета 1 типа характерно быстрое развитие симптомов

(в течение нескольких недель, месяцев).

Сахарный диабет 2 типа может длительное время себя не проявлять. Его можно выявить случайно, при очередном медицинском осмотре или сдаче анализов по поводу другого заболевания или при медицинском осмотре. Симптомы могут развиваться годами, быть мало выраженными.

Основные симптомы являются:

1. Полиурия – повышенное мочеотделение. По мере увеличения уровня глюкозы в крови, повышается ее содержание и в моче. Почки реагируют первыми и начинают выделять больше жидкости из организма, для того чтобы разбавить концентрацию глюкозы в моче. Характерно усиление мочеотделения в ночное время. В результате этого происходит усиленное мочевыделение (до 2 литров мочи в сутки).
2. Полидипсия – это неутолимая жажда, сухость во рту, что является следствием выделения большого количества жидкости из организма в виде мочи. Больные начинают много пить, чтобы утолить жажду и восполнить потерю жидкости с мочой.
3. Полифагия – это постоянное чувство голода. Это связано с нарушением обмена веществ, точнее сказать из – за того, что клетки не способны поглощать и перерабатывать глюкозу без помощи инсулина. Резкое снижение веса, похудение особенно характерны для больных с диабетом 1 типа. Это связано с повышенным разрушением белков и жиров из-за отсутствия глюкозы в энергетическом обмене клеток. Парадоксально то, что похудение развивается, несмотря на повышенный аппетит больного.

Симптомы, которые могут сопровождать диабет, но не являются обязательными:

* сухость кожных покровов;
* сухость во рту;
* мышечная слабость;
* гнойничковые образования и заболевания кожи;
* длительное и плохое заживление ран и рубцов;
* быстрая утомляемость;
* частые головные боли;
* зуд половых органов (чаще после мочеиспускания).

Выделяются три степени тяжести заболевания:

1. степень (легкая) – при этом повышение глюкозы в крови не превышает 8 ммоль/л натощак. При этом нет значительных колебаний глюкозы в течение суток, допустимы следы глюкозы в моче до 20 г/л. Возможны начальные проявления осложнений (ангионейропатии).
2. степень (средняя) – уровень глюкозы в крови натощак достигает 14 ммоль/л. Глюкоза в моче увеличивается до 40 г/л. Компенсация состояния достигается диетой и приемом сахароснижающих лекарственных препаратов. Возможны проявления осложнений.
3. степень (тяжелое течение) – уровень глюкозы натощак более 14 ммоль/л, в моче уровень глюкозы превышает 50 г/г. В этой стадии больные нуждаются в постоянной инсулинотерапии и ярко выражены сопутствующие осложнения.

Рассмотрим разновидности фаз компенсации сахарного диабета:

* Фаза компенсации, при которой больной чувствует себя хорошо, а с помощью терапии легко можно добиться нормальных цифр глюкозы в крови. В моче глюкозы не содержится.
* Фаза субкомпенсации. При данной фазе не удается снизить уровень глюкозы в крови ниже 13,9 ммоль/л. Появляется сахар в моче. Ацетона в моче не содержится.
* Фаза декомпенсации (самая тяжелая) – проводимая терапия не дает эффекта, и уровень сахара поднимается выше 14,0 ммоль/л. Количество глюкозы в моче увеличивается и появляется ацетон. Возможно развитие гипергликемической комы.

## 1.2 Корреляционно-регрессионный анализ

Корреляционно-регрессионный анализ является наиболее широко распространенным и гибким приемом обработки статистической информации.

Корреляционно-регрессионный анализ — это один из самых распространенных методов изучения отношений между численными величинами. Его основная цель состоит в нахождении зависимости между двумя параметрами и ее степени с последующим выведением   
уравнения. То есть, корреляционно-регрессионный анализ представляет из себя объединение методов корреляционного и регрессионного анализов.

Задачами корреляционно-регрессионного анализа являются:

* + установление типа уравнения регрессии;
  + определение параметров уравнения регрессии и оценка значимости параметров;
  + оценка тесноты и направления связи между переменными; - оценка значимости уравнения регрессии;
  + определение прогнозных значений зависимой переменной и оценка полученного прогноза.

Так как в корреляционно-регрессионном анализе используются методы корреляционного и регрессионного анализа, рассмотрим эти методы подробней.

Корреляционный анализ — раздел математической статистики, в котором изучаются задачи выявления статистических зависимостей между случайными величинами путем оценок различных коэффициентов корреляции. Методы корреляционного анализа дают хорошие результаты тогда, когда данные эксперимента можно считать выбранными из генеральной совокупности, распределенной по многомерному нормальному закону.

Невозможно управлять явлениями, предсказывать их развитие без изучения характера, силы и других особенностей связей. Поэтому методы исследования, изменения связей составляют чрезвычайно важную часть методологии научного исследования, в том числе и статистическую.

Связи между изучаемыми переменными подразделяются на функциональные и статистические. При функциональной связи определенному значению одной переменной величины соответствует строго определенное значение другой переменной.

При изменении одной из них на определенную величину, другая переменная изменяется на величину, в соответствии с видом функции, связывающей переменные.

Статистической называется связь между переменными или признаками, когда определенному значению факторного признака соответствует несколько различных значений результативного признака. Частным случаем статистической связи является корреляционная, которая проявляется в среднем, в массе наблюдений, как статистическая закономерность.

При корреляционной связи с изменением факторного признака на определенную величину изменяется среднее значение результативного признака. Обычно корреляционная зависимость представляется как функциональная зависимость между переменными в виде уравнения регрессии.

Корреляционной связью называют важнейший частный случай статистической связи, состоящий в том, что разным значениям одной переменной соответствуют различные средние значения другой. С изменением значения признака х закономерным образом изменяется среднее значение признака у; в то время как в каждом отдельном случае значение признака y (с различными вероятностями) может принимать множество различных значений.

Тесноту связи изучаемых явлений оценивает Коэффициент   
Пирсона (Kn).

Коэффициент Пирсона используется для изучения связи между двумя качественными признаками, каждый из которых состоит более чем из двух групп. Вычисляют по формуле:

,

где — показатель взаимной сопряженности:

,

где — объемы признака X по группам;

— объемы признака Y по группам;

— объемы выборок, относящихся к X и Y одновременно. [1.16]

Корреляционный коэффициент Пирсона может принимать значения в диапазоне .

По значению эмпирического корреляционного отношения судят о тесноте связи между признаками. Обычно придерживаются следующей шкалы:

— связь слабая;

— связь заметная;

— связь умеренно тесная;

— связь тесная;

— связь очень тесная.

После того как с помощью корреляционного анализа выявлено наличие статистических связей между переменными и оценена степень их тесноты, обычно переходят к математическому описанию зависимостей, то есть к регрессионному анализу.

Регрессионный анализ применяется в тех случаях, когда необходимо отыскать непосредственно вид зависимости х и у. При этом предполагается, что независимые факторы являются не случайными величинами, а результативный показатель у имеет постоянную, независимую от факторов дисперсию и стандартное отклонение.

Рассмотрим метод линейной регрессии.

Под линейностью имеется в виду, что переменная у предположительно находится под влиянием переменной х в зависимости

,

где b — постоянная величина (или свободный член уравнения);

a — коэффициент регрессии, определяющий наклон линии, вдоль которой рассеяны данные наблюдения. Это показатель, характеризующий изменение переменной yi при изменении значения xi на единицу. Если a > 0, переменные xi и yi положительно коррелированные, если a < 0 — отрицательно коррелированные;

εi — независимая нормально распределенная величина — остаток с нулевым математическим ожиданием и постоянной дисперсией:

,

где σ — среднее квадратическое отклонение. Отражает тот факт, что изменение yi будет неточно описываться изменением xi: присутствуют другие факты, не учтенные в данной модели.

Построение линейной регрессии сводится к оценке параметров a и b. Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений фактических значений результативного признака y от расчетных (теоретических) , минимальна:

,

где yi — фактические значения результативного признака y;

— расчетные значения результативного признака y.

То есть из всего множества линий линия регрессии на графике выбирается так, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была бы минимальной.

Таким образом, рассмотрено понятие корреляционно-регрессионного анализа, а также методы корреляционно-регрессионного анализа: нахождение корреляционной связи с помощью коэффициента Пирсона; построение парной линейной регрессии с помощью метода наименьших квадратов (МНК).

Рассмотрим применение корреляционно-регрессионного анализа на примере отрасли добычи полезных ископаемых в США.

# 2 ПРАКТИЧЕСКАЯ ЧАСТЬ

## 2.1 Корреляционно-регрессионный анализ признаков и рисков сахарного диабета для раннего прогнозирования.

В практической работе проведём корреляционно-регрессионный анализ Набор данных «Для прогнозирования риска сахарного диабета включает в себя следующие столбцы данных:

1. patient\_number – код (номер) пациента
2. cholesterol – холестерин
3. glucose – глюкоза
4. hdl\_chol – холестерин ЛПВП
5. age – возраст
6. height – рост
7. gender – пол
8. weight – вес
9. systolic\_bp – систолическое (верхнее) артериальное давление
10. diastolic\_bp – диастолическое (нижнее) артериальное давление
11. diabetes – есть диабет/нет диабета

Ниже приведен нормальный диапазон содержания различных растворенных веществ в крови:

* Уровень холестерина - менее 200
* Глюкоза 80-120 (в норме)
* Уровень холестерина ЛПВП более 60
* Систолическое АД - 120
* Диастолическое АД - 80

Ниже приведены моменты, вызывающие озабоченность в приведенном выше выводе кода:

Максимальный уровень холестерина, обнаруженный в данных, составляет 443, глюкозы 385, холестерина ЛПВП 120, систолического АД 250, диастолического 124

Минимальное значение холестерина ЛПВП - 12.

Вес 75% пациентов составляет от 99 до 200 килограммов.

На рисунке 2.0 представлена статистика

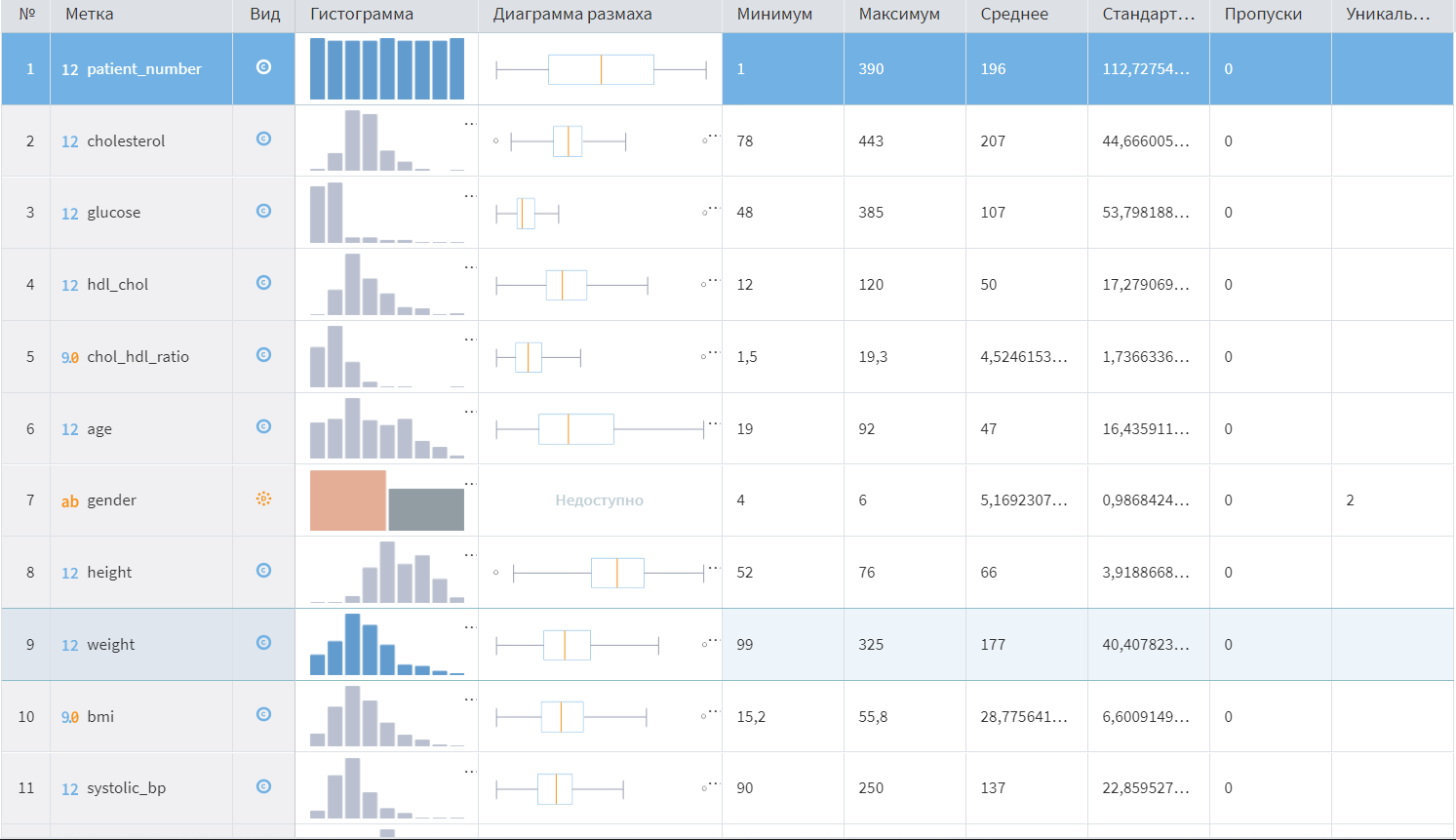


Рисунок 2.0 — Статистика

Импортируем наши наборы данных в сценарий пакета. Для этого в рабочую область сценария добавляем два узла «Excel файл» и для каждого в настройках указываем путь к нужному набору данных.

На Рисунке 2.1 представлено исходные данные

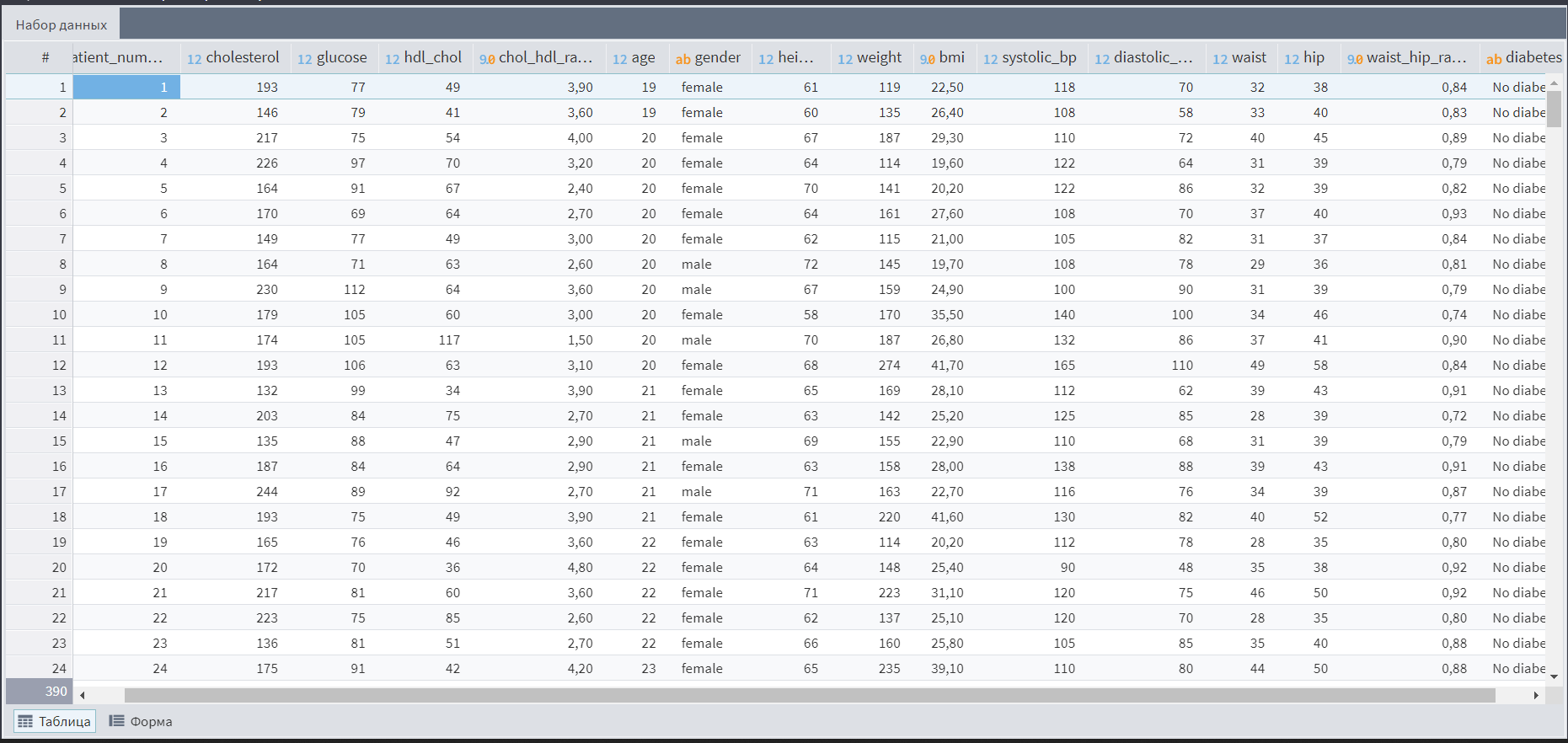


Рисунок 2.1 — Табличное представление исходных данных

Для проведения корреляционно-регрессионного анализа какой-либо отрасли требуется отфильтровать набор данных, выделив людей, у кого нет сахарного диабета и есть сахарный диабет. Нам требуется выделить столбец «diabetes». Воспользуемся узлом «Фильтр строк».

Визуализация результата в виде таблицы представлена на Рисунке 2.2.

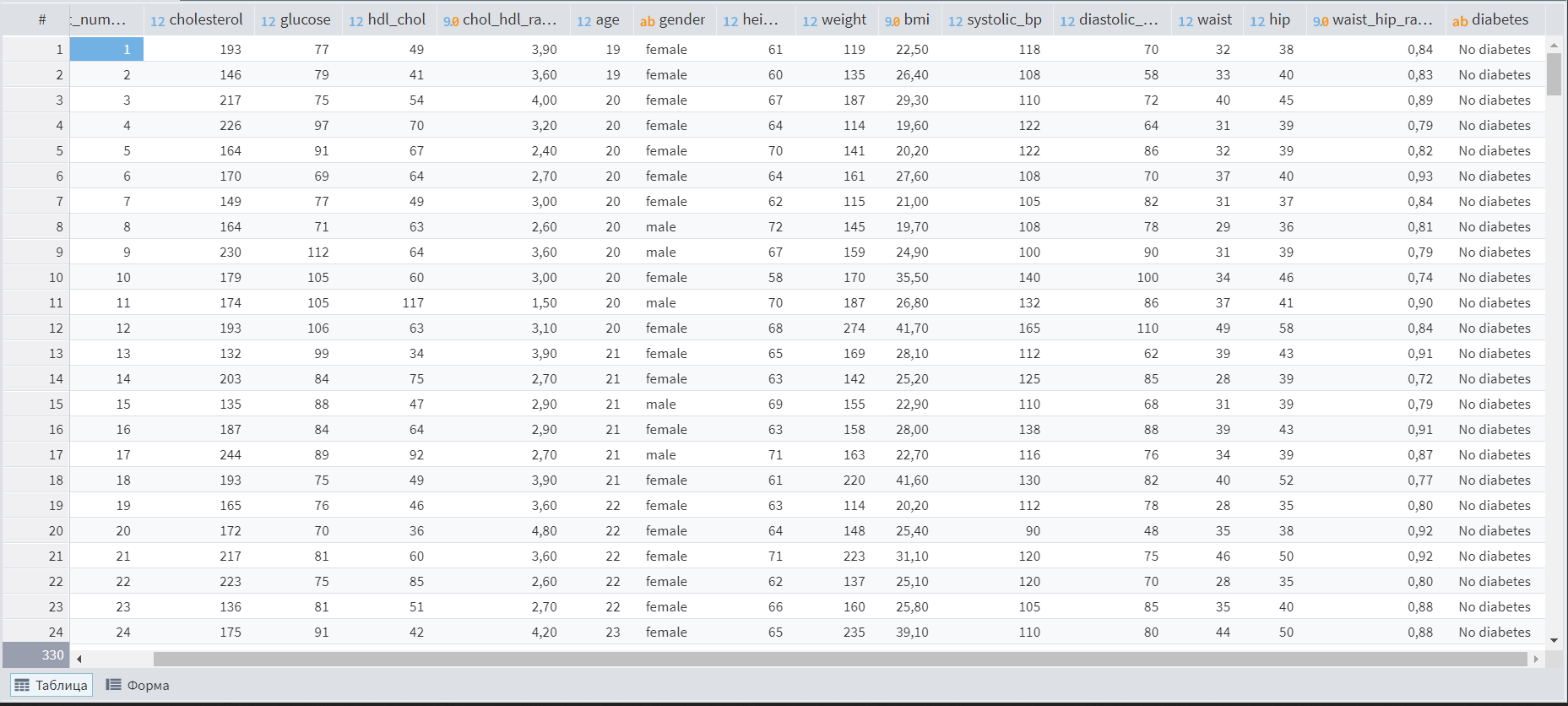


Рисунок 2.2 — Визуализация в виде таблицы набора данных с отфильтрованным набором данных, кого нет сахарного диабета

На Рисунке 2.3 показан отфильтрованный набор данных, выделив людей, у кого есть сахарный диабет.

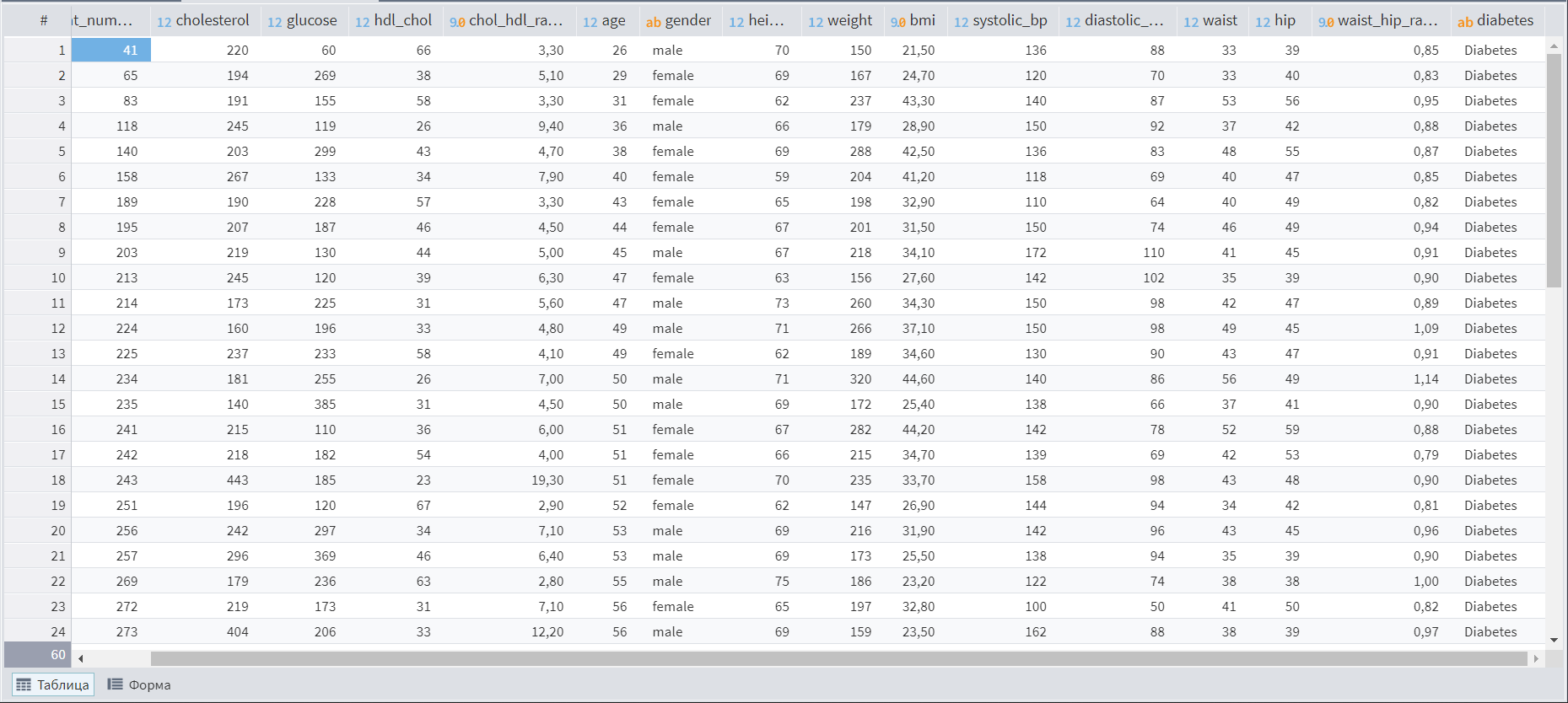
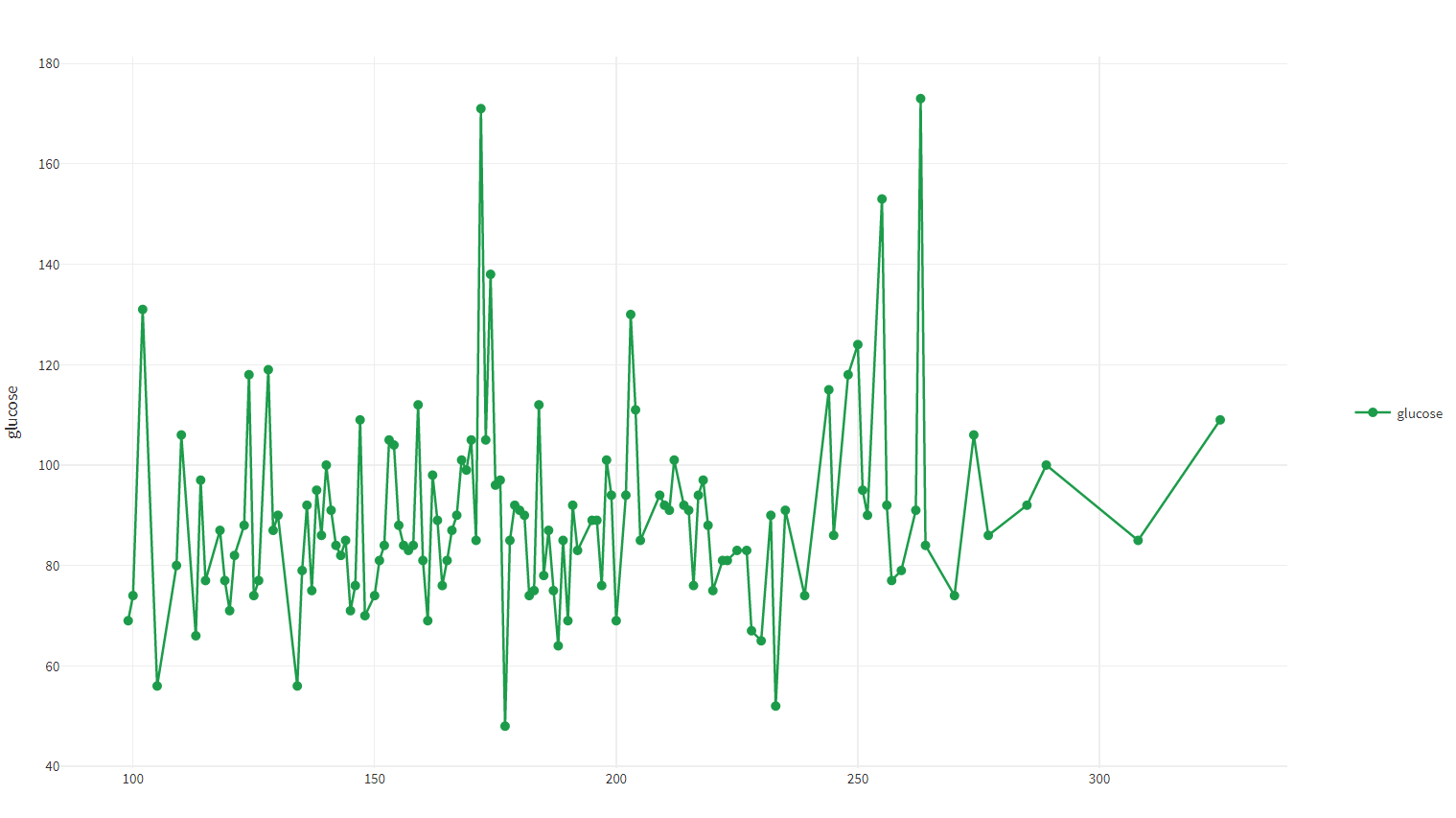


Рисунок 2.3 — Визуализация в виде таблицы набора данных с отфильтрованным набором данных, кого есть сахарного диабета

Общепризнано, что повышенный вес и пожилой возраст являются двумя основными факторами, вызывающими диабет.

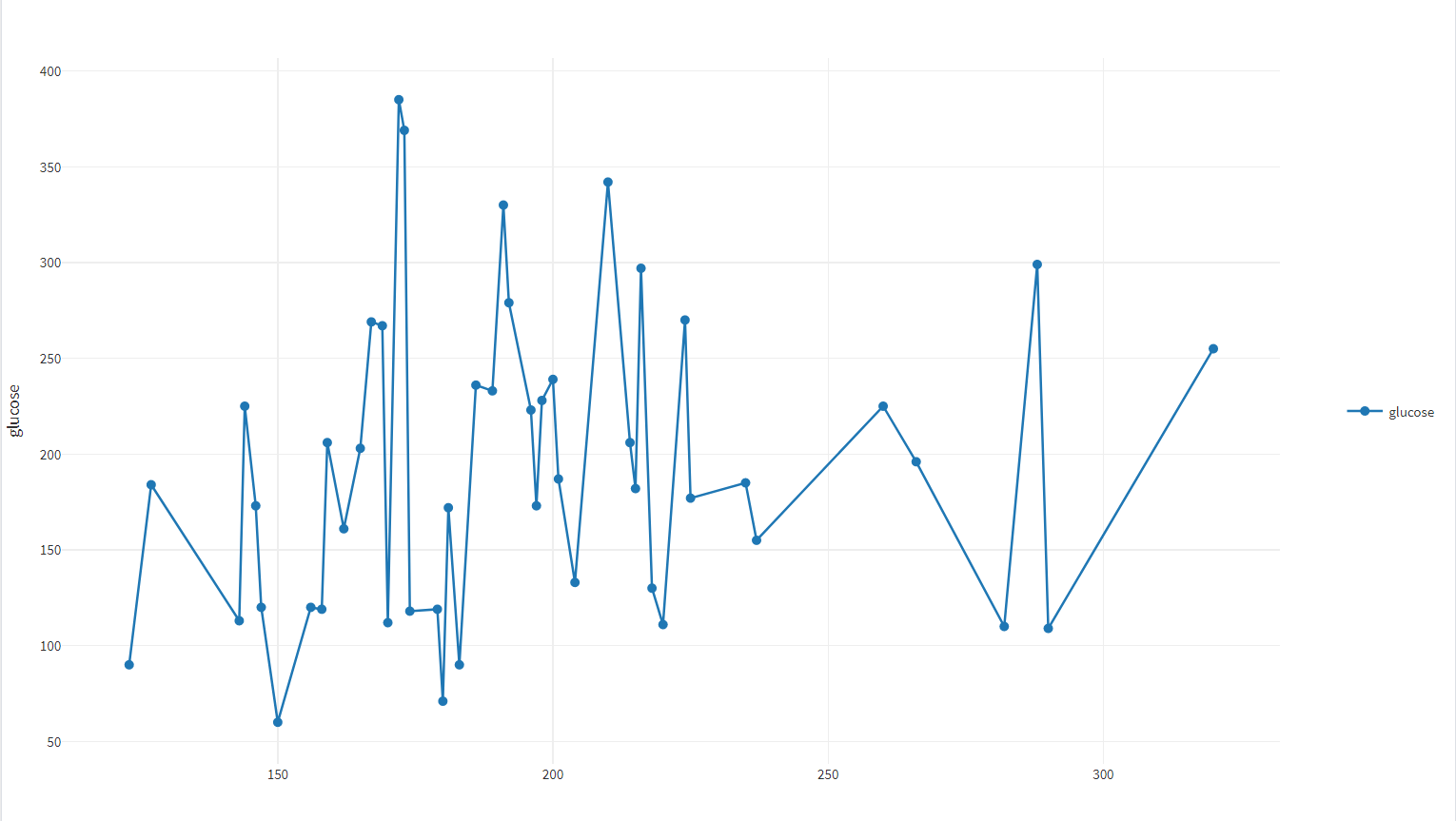
Пациенты с сахарным диабетом имеют более высокий уровень глюкозы и больший вес по сравнению с недиабетическими пациентами.

На Рисунке 2.4 показан график зависимости «glucose» от «weight», у которого нет сахарного диабета.



**Рисунок 2.4** — **График зависимости «glucose» от «weight»**

На Рисунке 2.5 показан график зависимости «glucose» от «weight», у которого есть сахарный диабет.

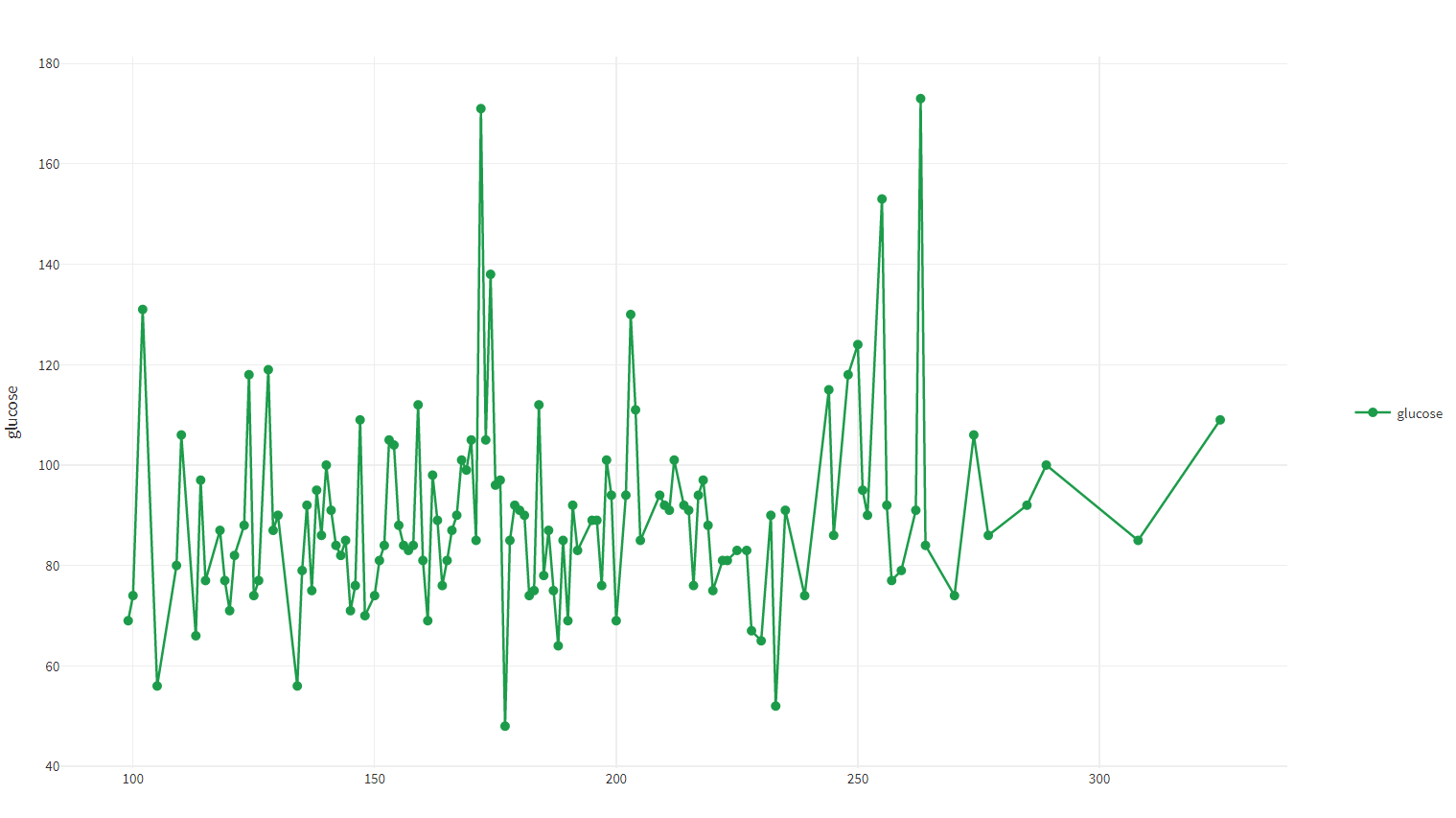


**Рисунок 2.5** — **График зависимости «glucose» от «weight»**

Требуется отфильтровать набор данных, выделив отдельно мужчин и отдельно женщин.

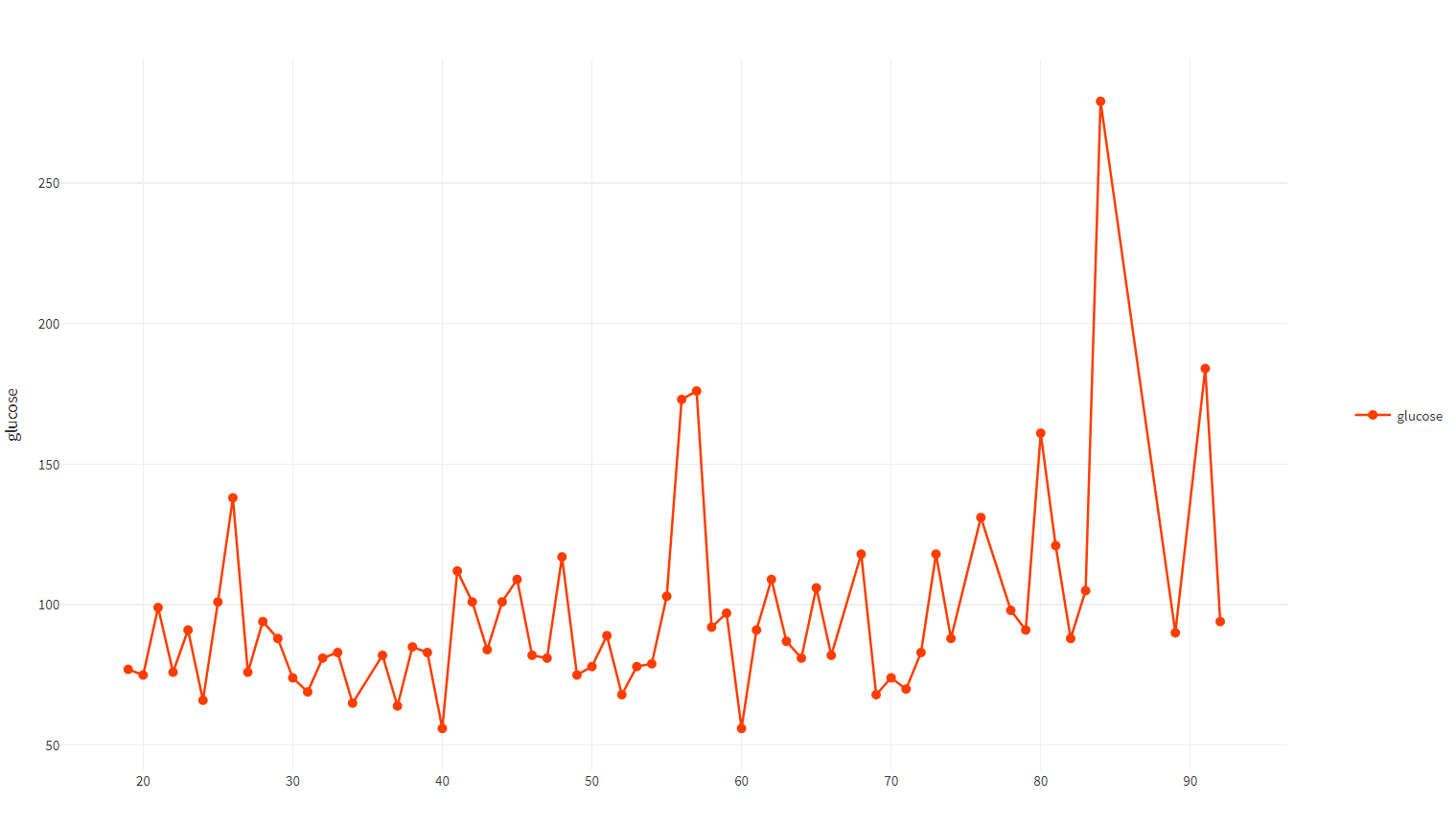
Возраст напрямую не определяется, но более высокий уровень глюкозы у пожилых людей может быть причиной развития у них диабета, также у мужчин в возрасте от 40 до 80 лет уровень глюкозы в крови выше, чем у женщин.

На Рисунке 2.6 показан график зависимости «glucose» от «age», глюкозы от возраста мужчин.



**Рисунок 2.6** — **График зависимости «glucose» от «age»**

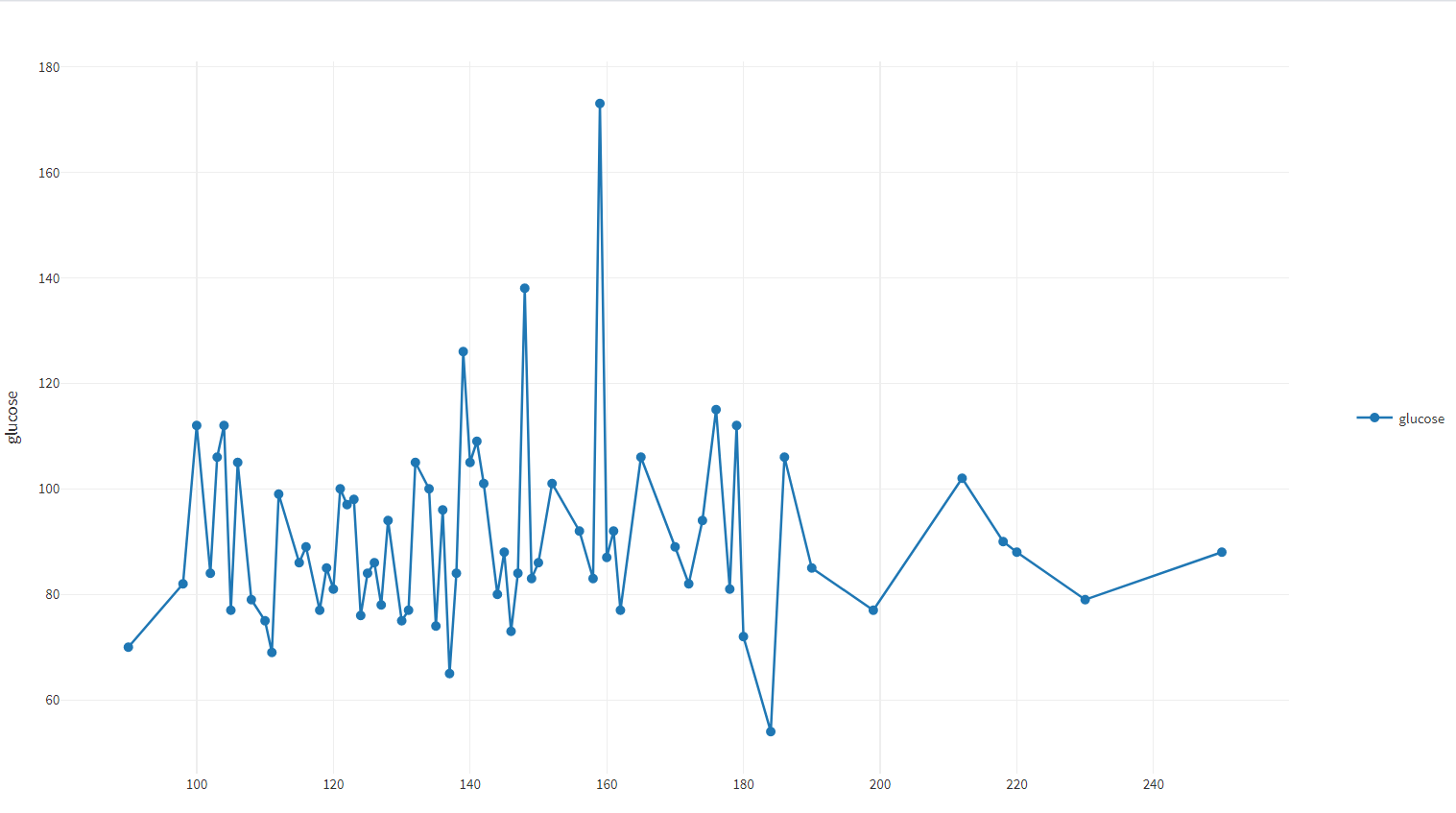
На Рисунке 2.7 показан график зависимости «glucose» от «age», глюкозы от возраста женщин.



**Рисунок 2.7** — **График зависимости «glucose» от «age»**

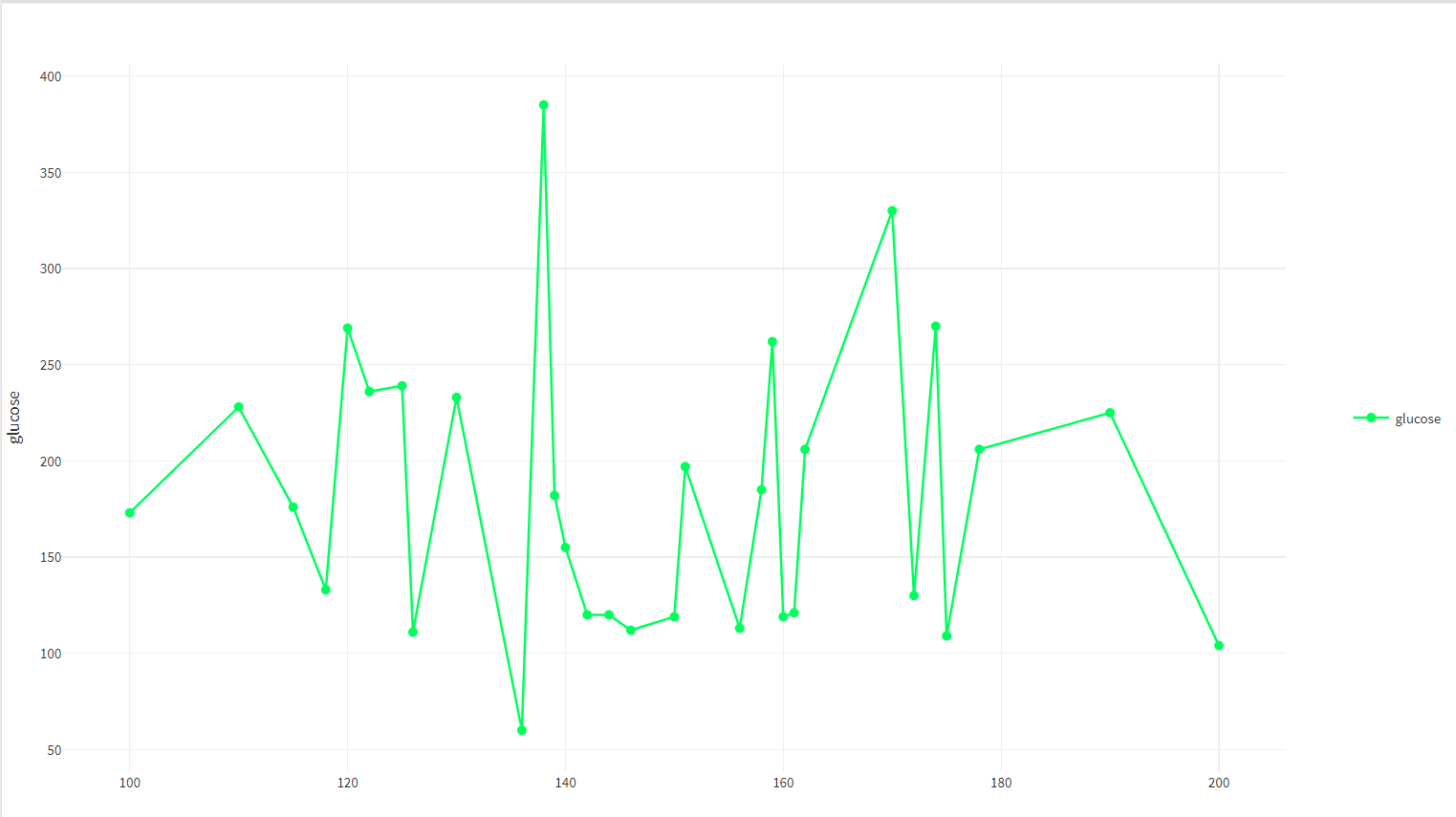
Артериальное давление напрямую не связано с сахарным диабетом, поскольку пациенты с самым высоким артериальным давлением считаются недиабетическими.

На Рисунке 2.8 показан график зависимости «glucose» от «systolic\_hp», у которых нет сахарного диабета.



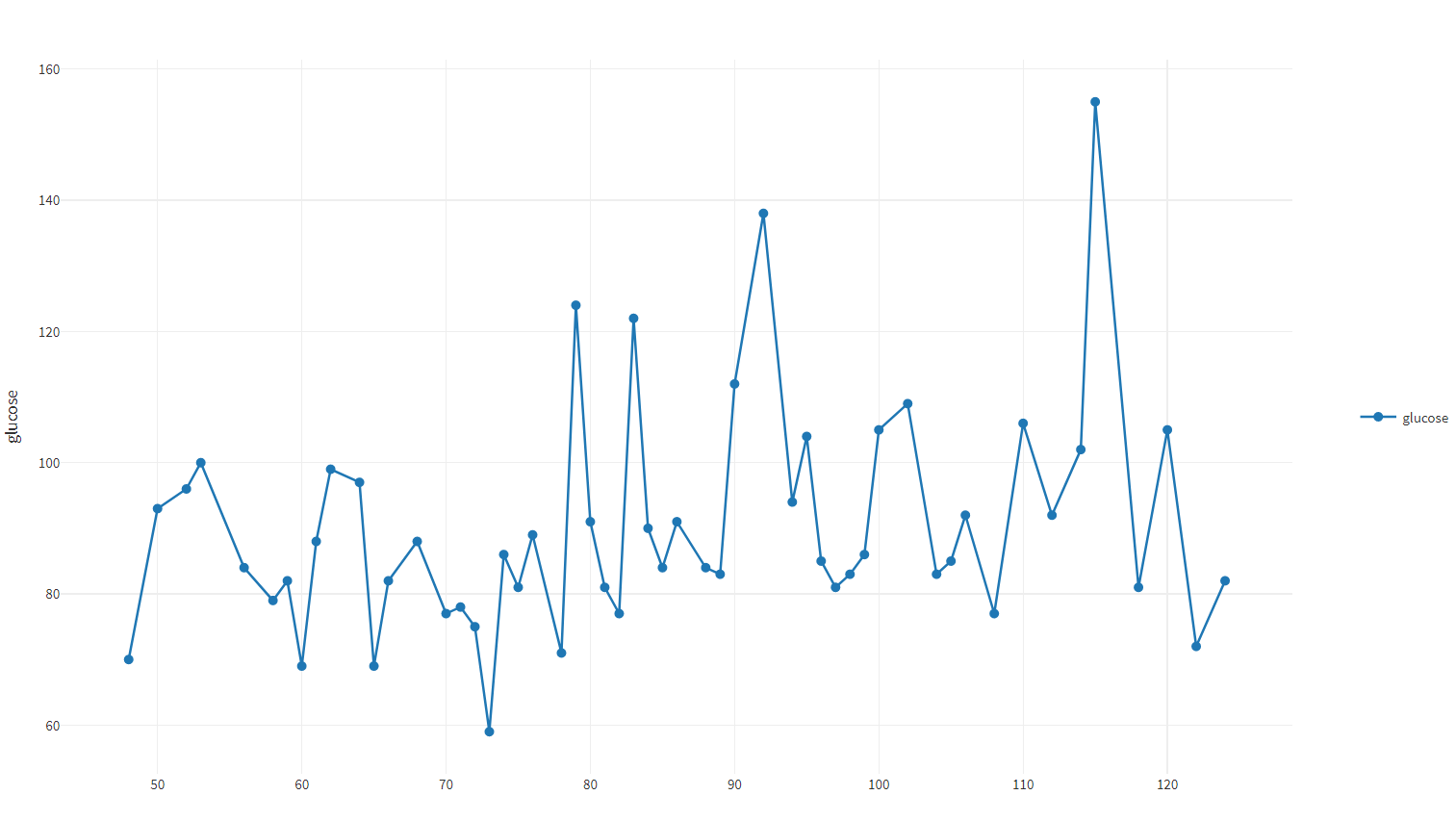
**Рисунок 2.8** — **График зависимости «glucose» от «systolic\_hp»**

На Рисунке 2.9 показан график зависимости «glucose» от «systolic\_hp», у которых есть сахарный диабет.



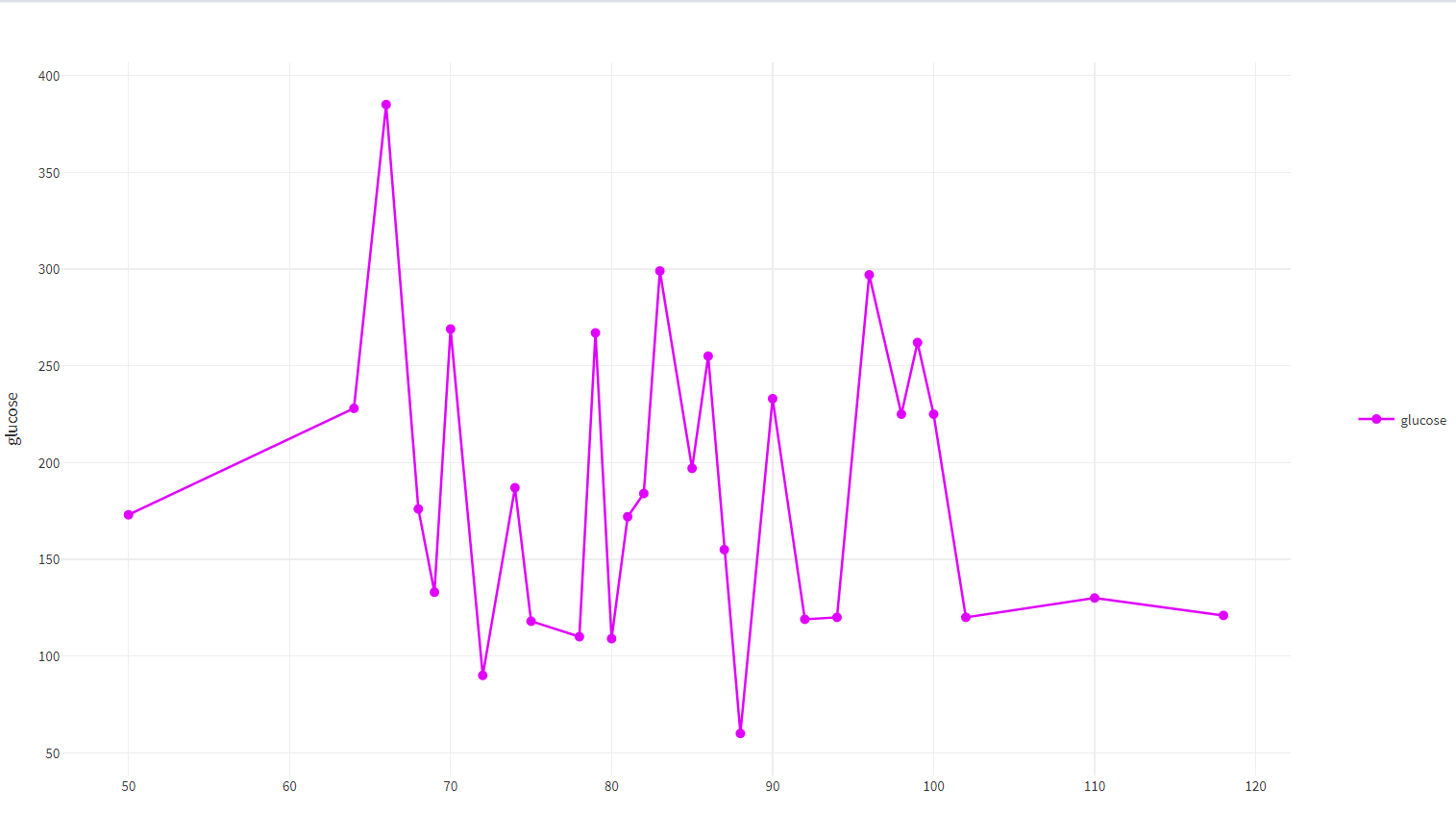
**Рисунок 2.9** — **График зависимости «glucose» от «systolic\_hp»**

На Рисунке 2.10 показан график зависимости «glucose» от «diastolic\_bp», у которых нет сахарного диабета.



**Рисунок 2.10** — **График зависимости «glucose» от «diastolic\_bp»**

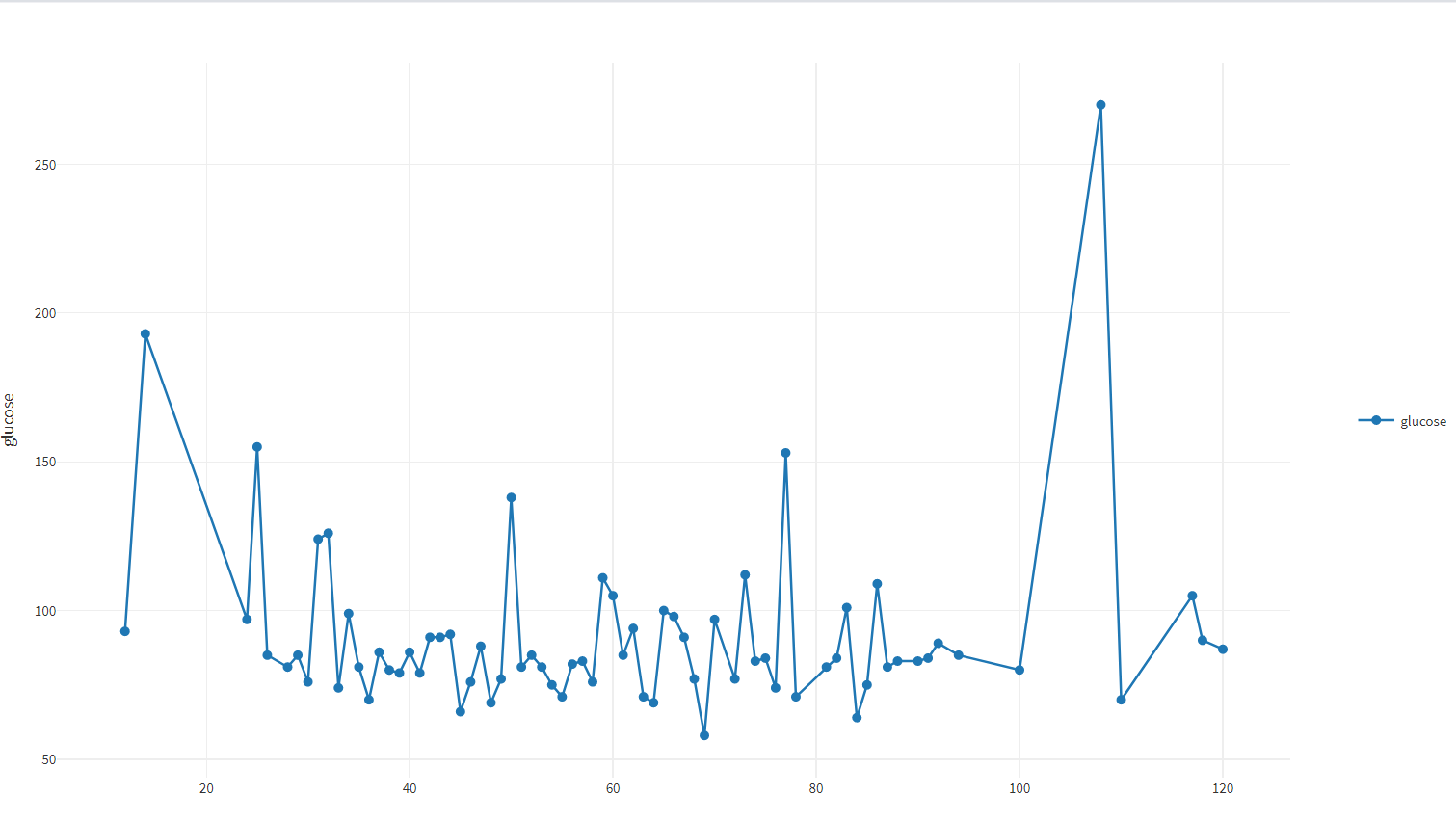
На Рисунке 2.11 показан график зависимости «glucose» от «diastolic\_bp», у которых есть сахарный диабет.



**Рисунок 2.11** — **График зависимости «glucose» от «diastolic\_bp»**

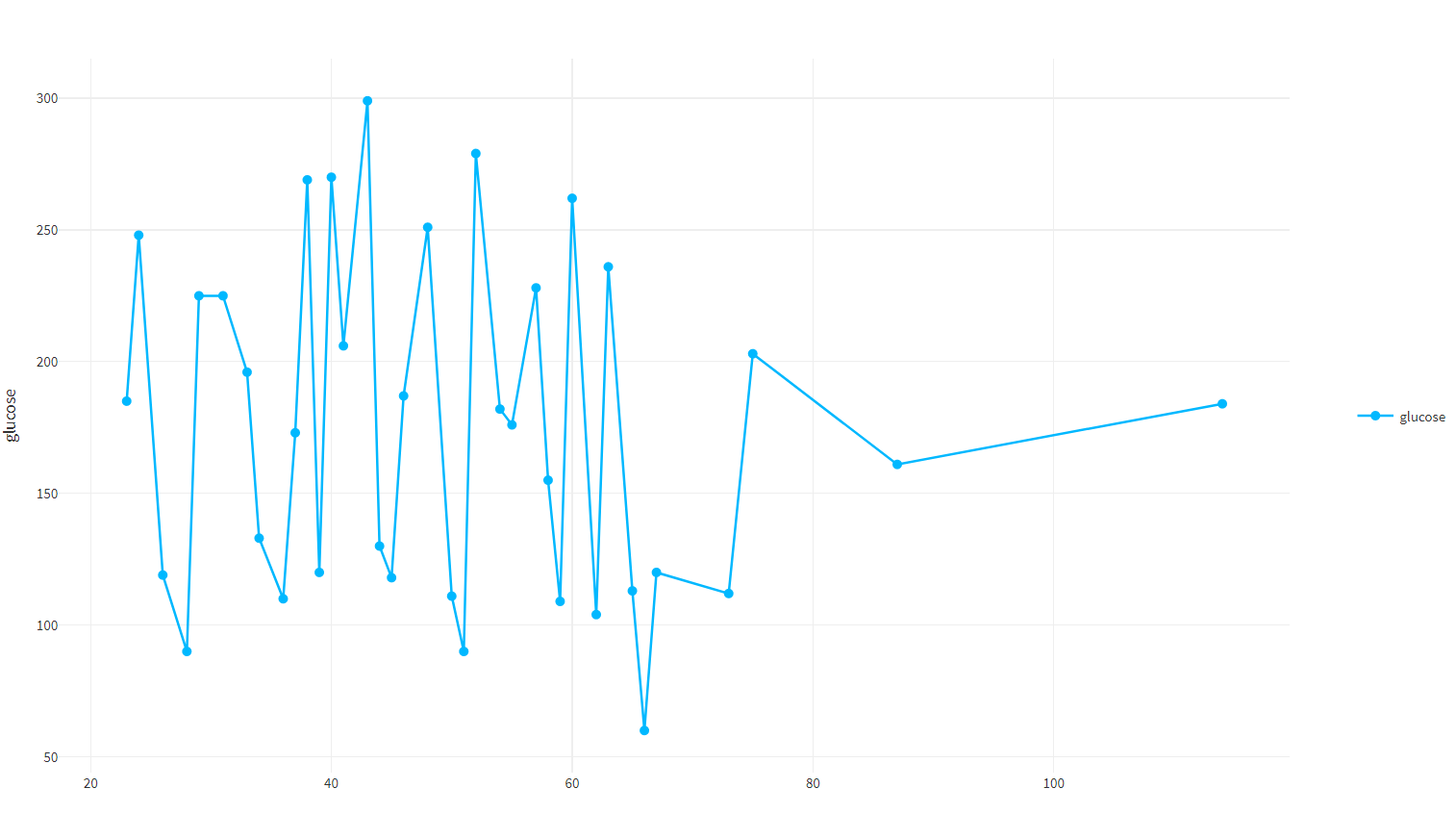
Пациенты с сахарным диабетом имеют более низкий уровень холестерина ЛПВП.

На Рисунке 2.12 показан график зависимости «glucose» от «hdl\_chol», у которых нет сахарного диабета.



**Рисунок 2.12** — **График зависимости «glucose» от «hdl\_chol»**

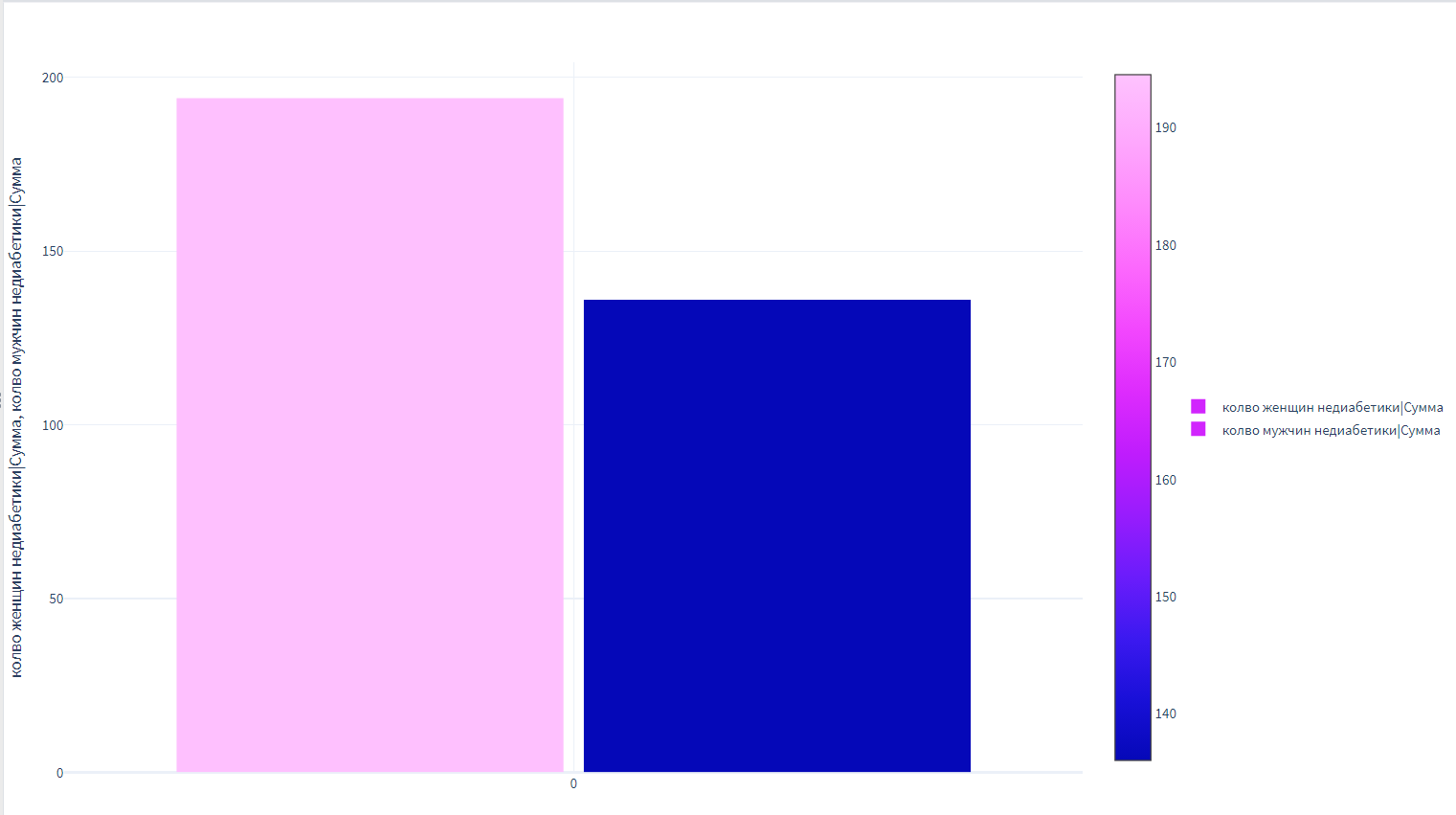
На Рисунке 2.13 показан график зависимости «glucose» от «hdl\_chol», у которых есть сахарный диабет.



**Рисунок 2.13** — **График зависимости «glucose» от «diastolic\_bp»**

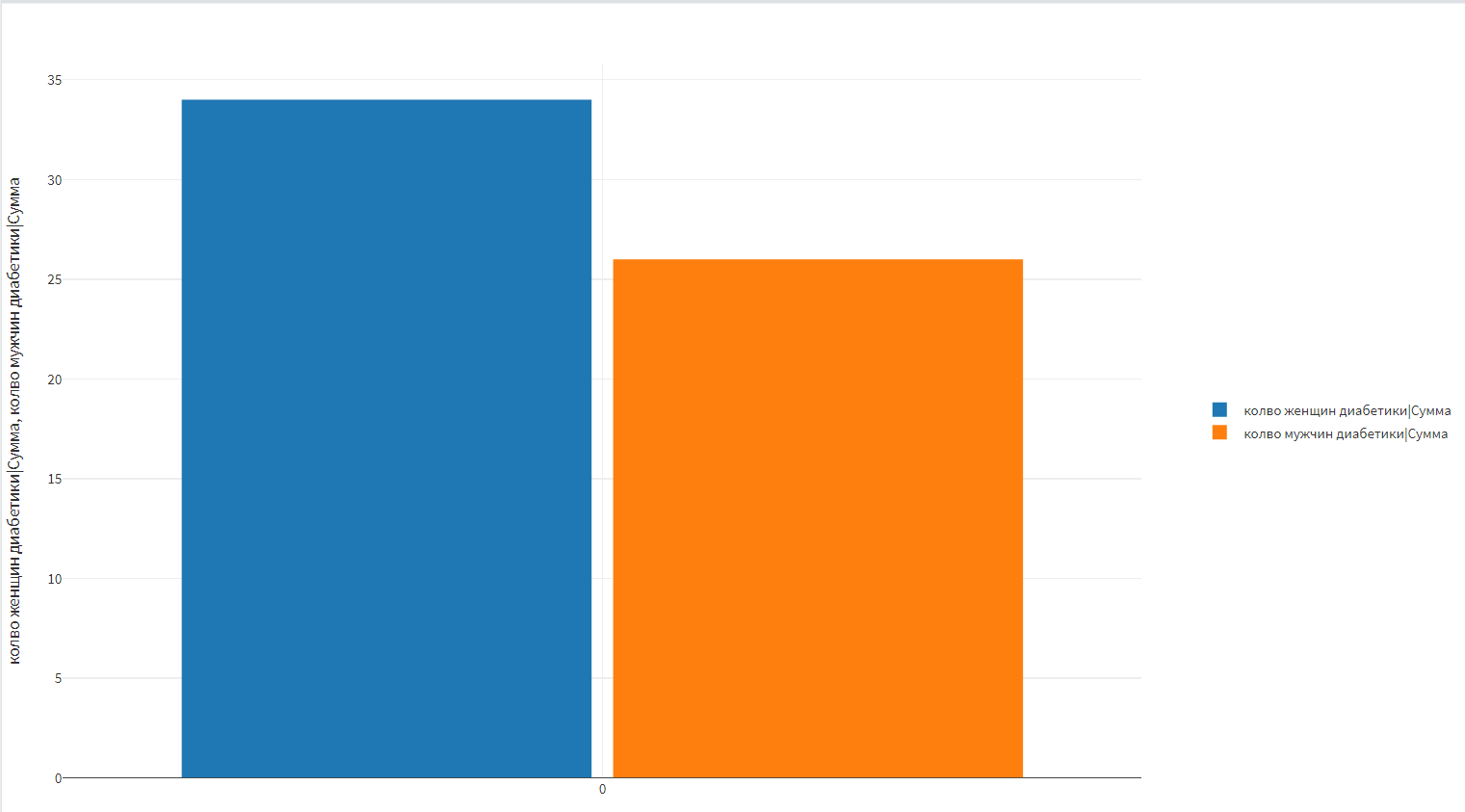
Женщин, страдающих диабетом, больше, чем мужчин, страдающих диабетом.

На Рисунке 2.14 показана столбчатая диаграмма.



**Рисунок 2.14** — **Сравнение между женщинами и мужчинами, которые не диабетики**

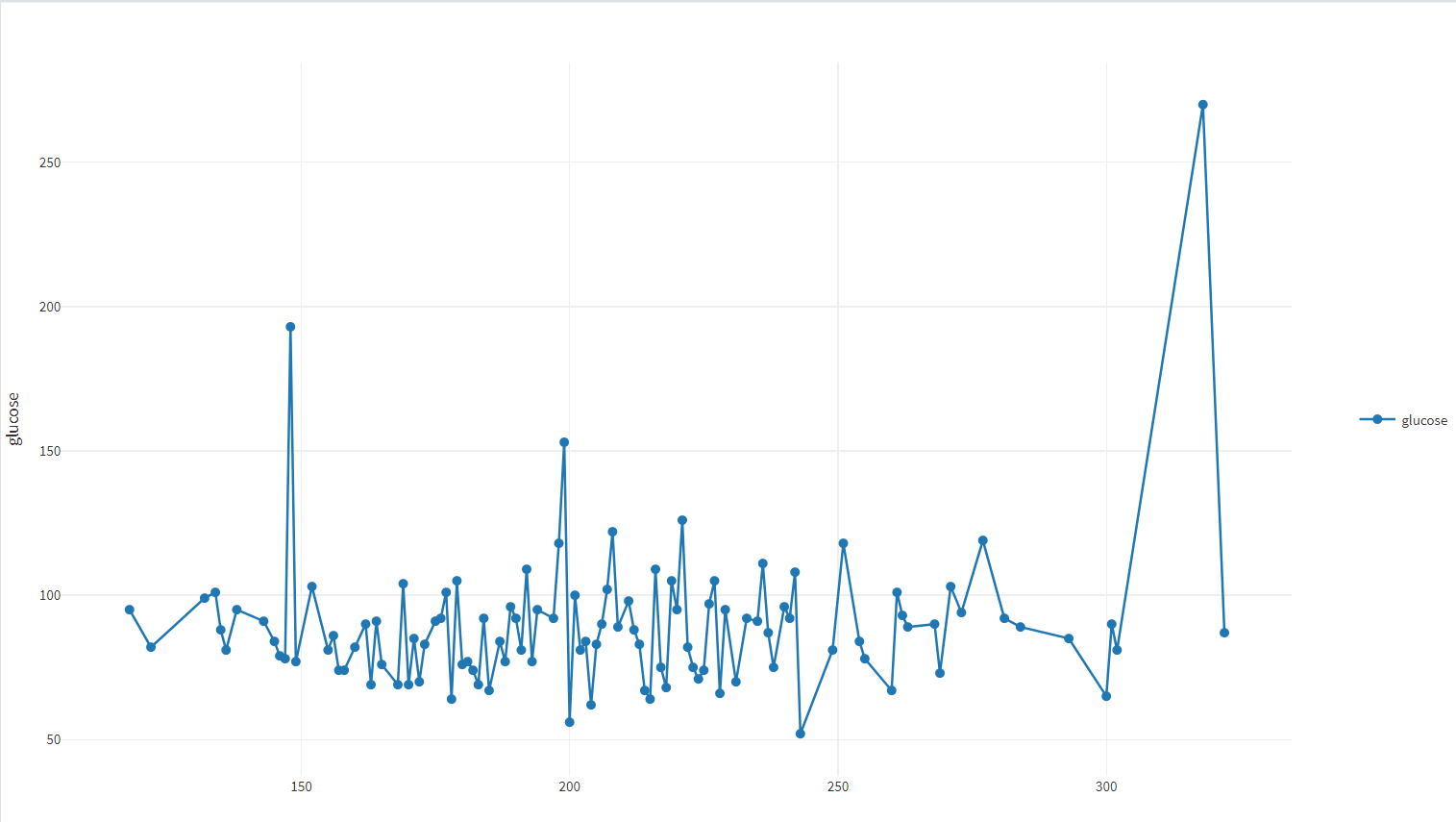
На Рисунке 2.15 показана столбчатая диаграмма.



**Рисунок 2.15** — **Сравнение между женщинами и мужчинами, которые диабетики**

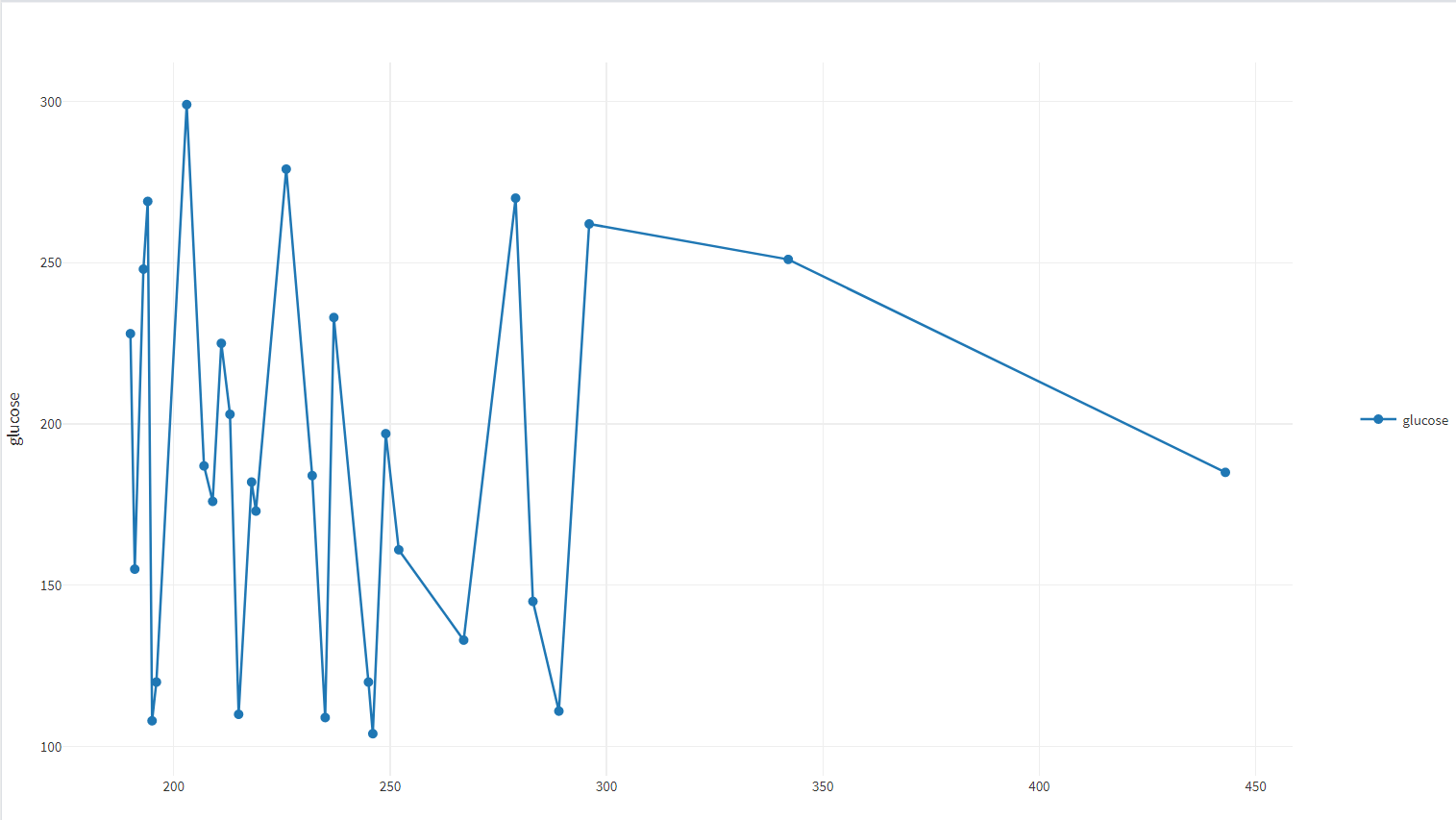
Более высокий уровень холестерина наблюдается у пациентов с сахарным диабетом

На Рисунке 2.12 показан график зависимости «glucose» от «cholesterol», у которых нет сахарного диабета.



**Рисунок 2.12** — **График зависимости «glucose» от «cholesterol»**

На Рисунке 2.13 показан график зависимости «glucose» от «cholesterol», у которых есть сахарного диабета.



**Рисунок 2.13** — **График зависимости «glucose» от «cholesterol»**

Как мы видели на приведенных выше графиках, глюкоза напрямую связана с причиной диабета. Хороший холестерин или ЛПВП- холестерин отрицательно связан, вес, возраст и холестерин также влияют, но в очень меньшей степени, у пациентов мужского пола уровень глюкозы выше, но из общего числа пациентов у женщин было больше случаев диабета, чем у мужчин, но разница была очень небольшой.

Для проведения корреляционно-регрессионного анализа необходимо разделить показатели на независимые (x) и зависимые (y) переменные. В данном случае зависимой переменной будет показатель «Люди, болеющие диабетом», а независимыми — показатели «Холестерин», «Глюкоза» и «холестерин ЛПВП». Столбцы «Код пациента» в корреляционно-регрессионном анализе не участвуют.

Сначала нужно найти значение коэффициента Пирсона и по нему определить тесноту связи между зависимой и независимыми переменными. Для вычисления коэффициента Пирсона используем узел «Корреляционный анализ». Результат представлен на Рисунке 2.14.

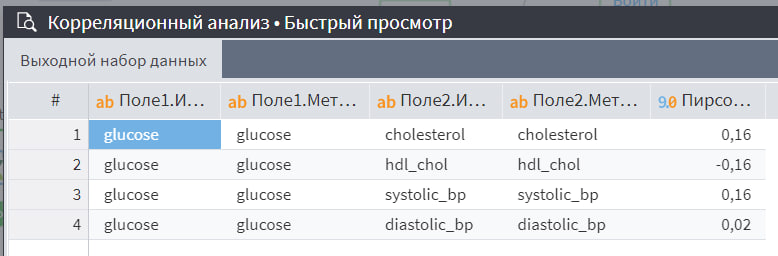


Рисунок 2.14 — Результат вычисления коэффициента Пирсона

Из Рисунка 2.14 мы видим, что коэффициент корреляции Пирсона между показателем «Глюкозой» и показателем «Диабетиков» равен -0,16;

Коэффициент корреляции Пирсона же между показателем «Холестерином» и показателем «Диабетиков» равен -0,20.

Более высокий уровень холестерина наблюдается у пациентов с сахарным диабетом

Посмотрим на сравнение значений коэффициента Пирсона всех независимых переменных. Визуализация представлена на Рисунке 2.12.

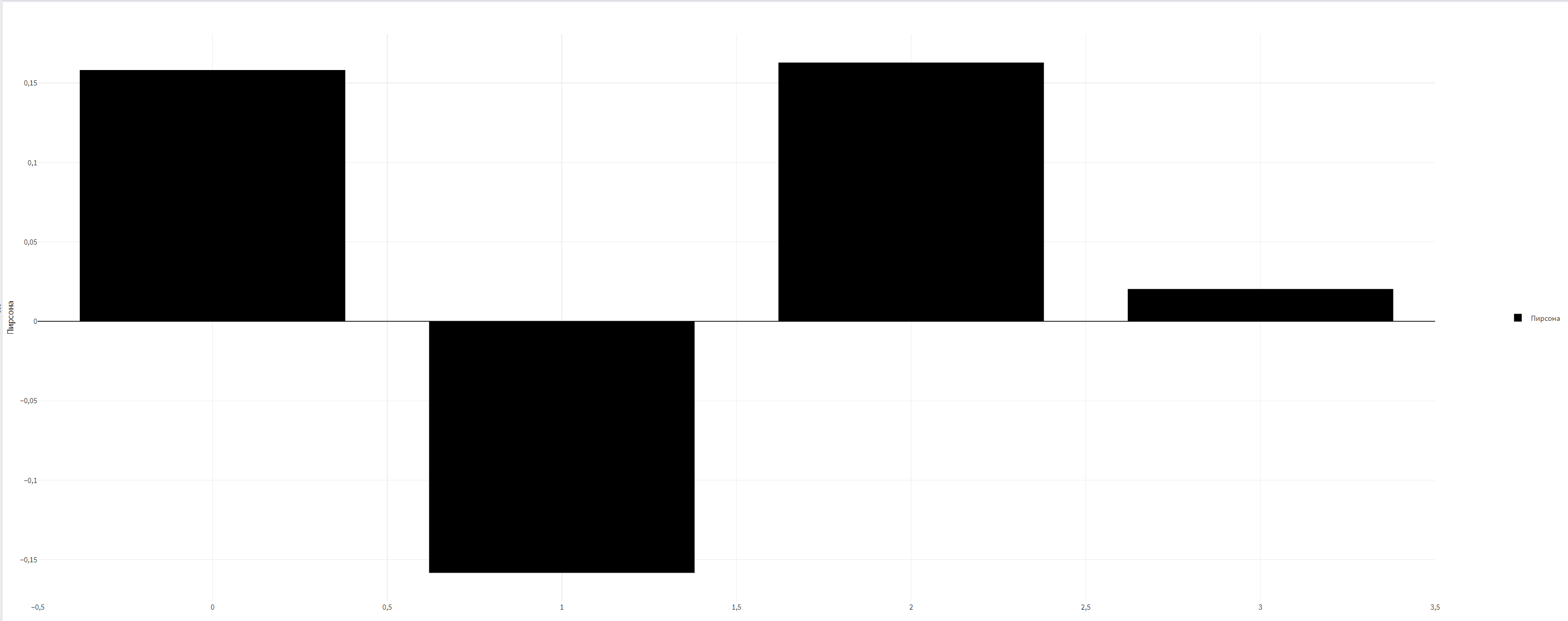


Рисунок 2.12 — Сравнение коэффициентов Пирсона

После анализа AUC можем заметить, что у нас «отличное» качество. Значение = 0,93.

На рисунке 2.13 мы видим результаты AUC - анализа.

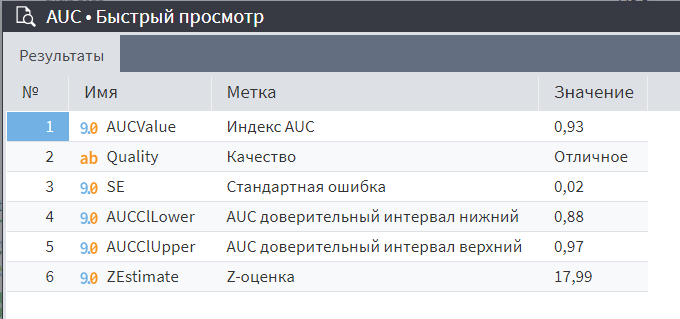


Рисунок 2.13 — Результат AUC-анализа

Для проведения анализа необходимо воспользоваться узлом ROC – анализ зависимой переменной будет показатель «Чувствительность», на показатели «1 - специфичность».

На рисунке 2.14 мы видим график похожую на логистическую регрессию.

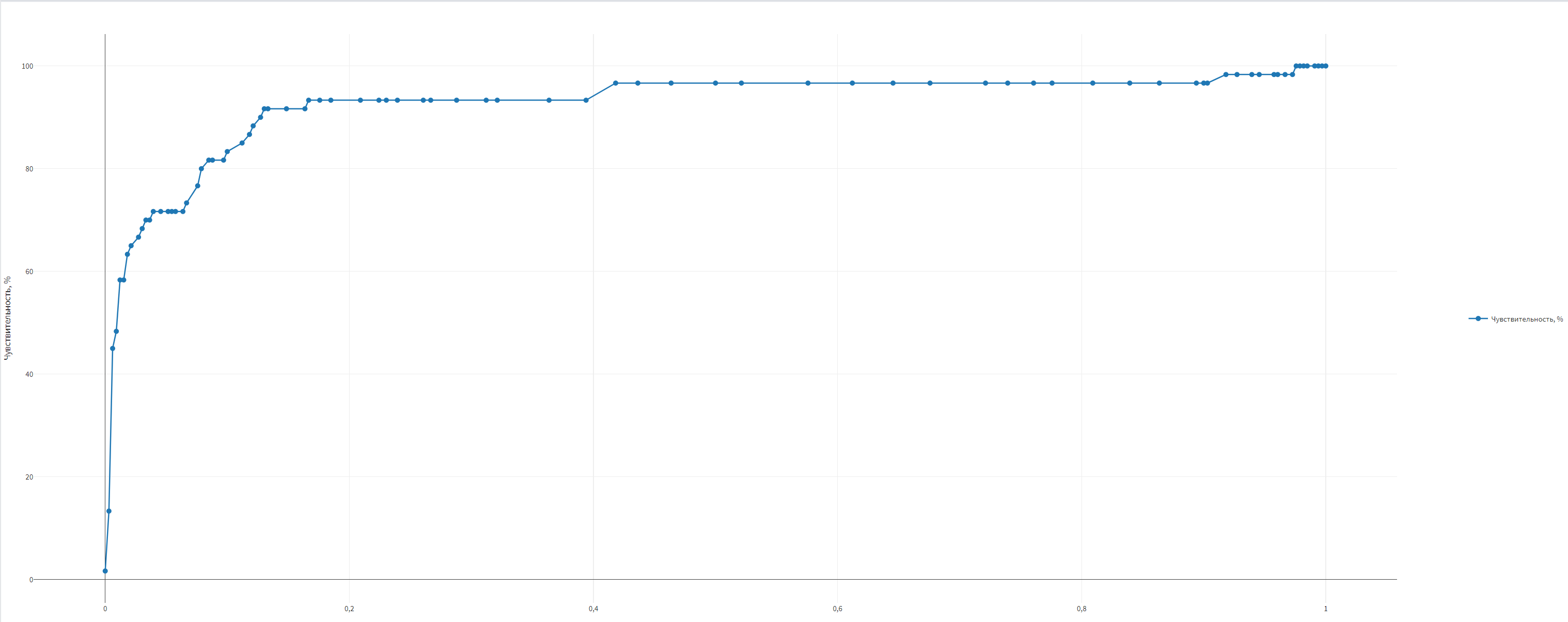


Рисунок 2.14 — Логистическая регрессия

# ЗАКЛЮЧЕНИЕ

Сахарный диабет – это нарушение обмена веществ, характеризующееся повышением содержания сахара в крови.

Заболевание возникает в результате дефектов выработки инсулина, дефекта действия инсулина или обоих этих факторов. Помимо повышенного уровня сахара крови, заболевание проявляется выделением сахара с мочой, обильным мочеиспусканием, повышенной жаждой, нарушениями жирового, белкового и минерального обменов и развитием осложнений.

Цель данной курсовой работы — провести корреляционно-регрессионный анализ признаков и рисков сахарного диабета для раннего прогнозирования.

В ходе выполнения данной курсовой работы проведен корреляционно-регрессионный анализ влияния выпуска валовой продукции в отрасли на количество работников в отрасли с использованием low-code платформы Loginom.

Задачи, выполненные в данной курсовой работе:

* + изучены научная и методическая литературы о заболеваниях сахарным диабетом и корреляционно-регрессионном анализе;
  + выполнен корреляционно-регрессионный анализ собранных данных;
  + использованы знания математической статистики с использованием современных средств обработки данных: аналитической платформы Loginom;
  + пройдено обучение оформлению официальных документов.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

## Теоретическая часть

* 1. Сахарный диабет / Свободная энциклопедия Wikipedia [Электронный ресурс]. <https://ru.wikipedia.org/wiki/Сахарный_диабет>
  2. Корреляционно-регрессионный анализ / Научный словарь-справочник «Справочник24» [Электронный ресурс] <https://spravochnick.ru/ekonomicheskiy_analiz/korrelyacionno-regressionnyy_analiz/>
  3. Корреляционно-регрессионный анализ: пример, задачи, применение. Метод корреляционно-регрессионного анализа / Интернет-портал BusinessMan.ru [Электронный ресурс]. https://businessman.ru/new-korrelyacionno-regressionnyj-analiz-primer-zadachi-primenenie.html
  4. П.С. Бондаренко, Г.В. Горелова, И.А. Кацко, А.Е. Жминько, Т.В. Соловьёва, С.А. Кацко, С.К. Куижева, А.А. Митус, Н.Б. Паклин, А.Е. Сенникова. Эконометрика. Практикум: учебно-практическое пособие / коллектив авторов; под ред. И.А. Кацко. — Москва: КНОРУС, 2019. — 218 с.
  5. Александрова, О.В. Статистические методы решения технологических задач: учебное пособие / О.В. Александрова, Т.А. Мацеевич, Л.В. Кирьянова, В.Г. Соловьев. – Москва: Издательство МИСИ – МГСУ, 2017. — 154 с.
  6. Круценюк, К.Ю. Корреляционно-регрессионный анализ в эконометрических моделях: учебное пособие / К. Ю. Круценюк. – Норильск: НГИИ, 2018. — 108 с.
  7. Ганичева, А. В. Прикладная статистика: учебное пособие для спо / А. В. Ганичева. — Санкт-Петербург: Лань, 2021. — 164 с.
  8. Полякова, В.В. Основы теории статистики: [учеб. пособие] / В. В. Полякова, Н. В. Шаброва. – Екатеринбург: Изд-во Урал. ун-та, 2016. — 148 с.
  9. Метод корреляционно-регрессионного анализа / Студенческие реферативные статьи и материалы "Studref" [Электронный ресурс]. <https://studref.com/591347/ekonomika/metod_korrelyatsionno_regressionnogo_analiza>
  10. Кийко, П. В. Эконометрика. Регрессионные модели: учебное пособие / П. В. Кийко, Н. В. Щукина. — Омск: Омский ГАУ, 2021. — 83 с.
  11. Герасимова, Е. А. Эконометрика: регрессионный анализ: учебно-методическое пособие / Е. А. Герасимова, М. Ю. Карышев. — Самара: СамГУПС, 2016. — 98 с.

## Практическая часть

* 1. Loginom / Документация [Электронный ресурс]. https://help.loginom.ru/userguide/

# ПРИЛОЖЕНИЯ

Приложение А — Графический материал.

## Приложение А

На рисунке 3.1 представлен сценарий проекта в аналитической платформе Loginom, полученный в результате выполнения данной курсовой работы.

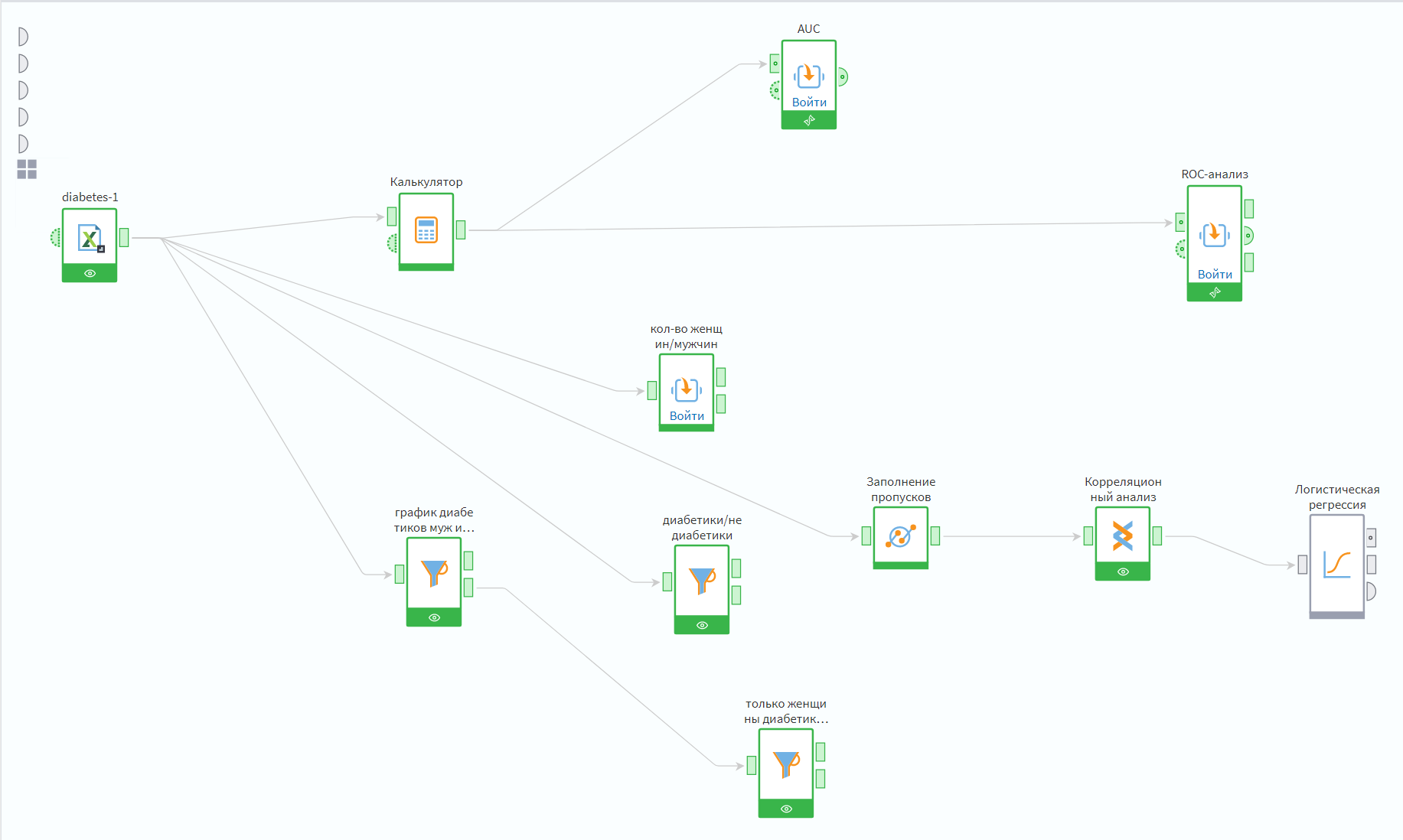


Рисунок 3.1 — Сценарий проекта в аналитической платформе Loginom

Этапы выполнения сценария:

1. Импорт исходных данных (Excel файлы).
2. Предобработка исходных данных.
3. Корреляционно-регрессионный анализ.