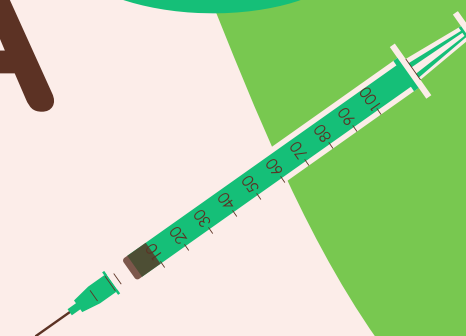
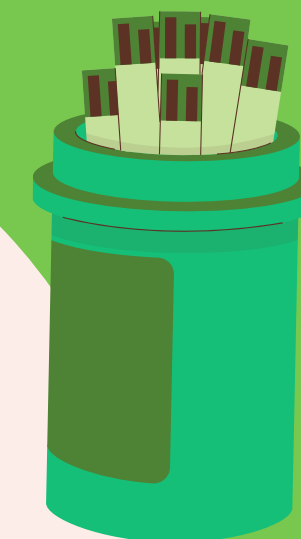
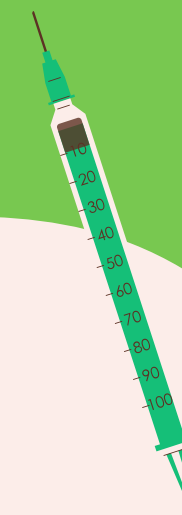
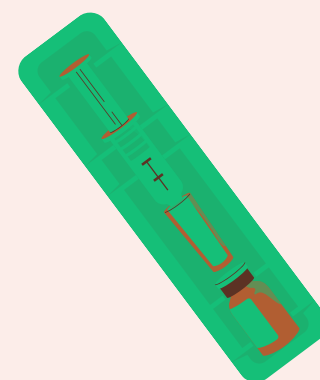
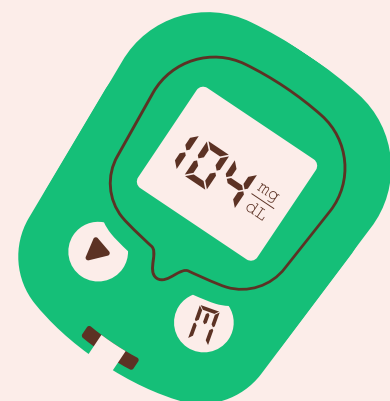


ПРЕДСКАЗАНИЕ РИСКА РАЗВИТИЯ ДИАБЕТА

Чубара Дарья ІА-213



ПРОБЛЕМАТИКА ТЕМЫ

Диабет стал глобальной проблемой из-за роста заболеваемости (по всему миру диабетом болеет около 463 миллионов человек), ограниченного доступа к лечению и недостаточной осведомленности.



ЦЕЛИ И ЗАДАЧИ

Цели: Предсказание диабета, выявление рисков, исследование факторов.

Задачи: Подготовка данных, выбор и оптимизация моделей, анализ зависимостей.



ДИЗАЙН ЭКСПЕРИМЕНТА

Предикторы:

- уровень сахара
- уровень глюкозы
- вес
- пол
- возраст
- гипертония
- сердечные заболевания

Методы анализа:

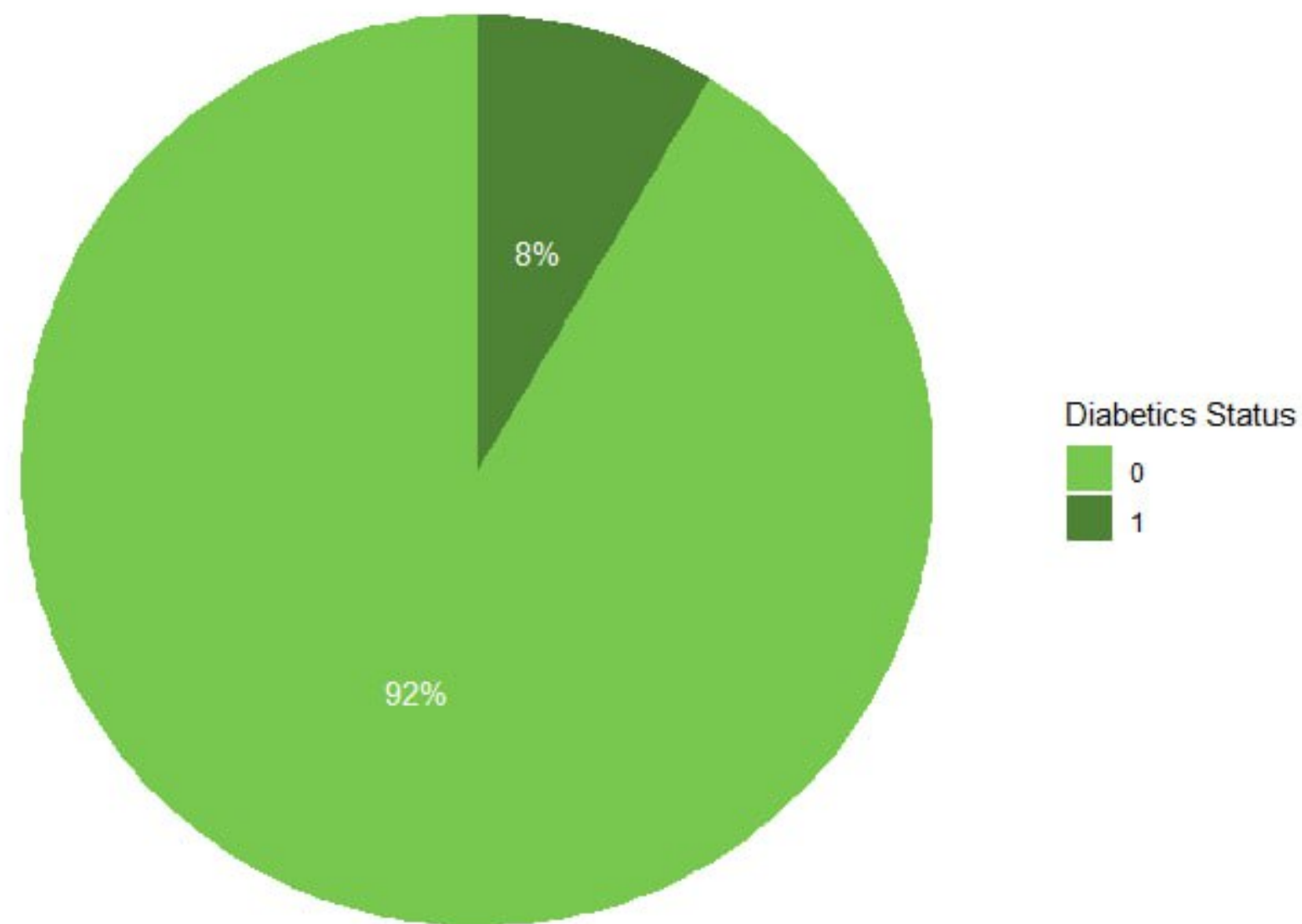
- логистическая регрессия

Метрики:

- confusion matrix
- accuracy
- precision
- recall

СКОЛЬКО БОЛЬНЫХ?

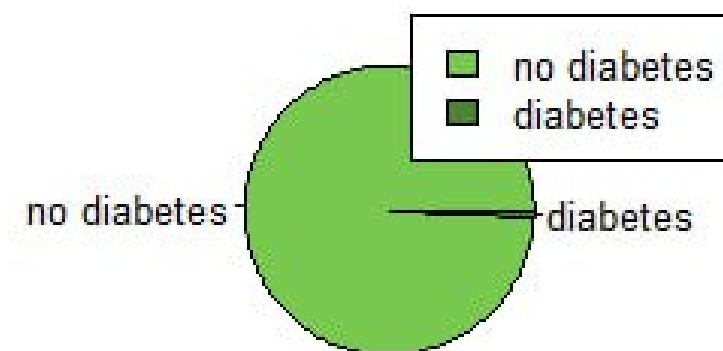
Distribution of Diabetes Cases



Люди, болеющие диабетом, составляют всего лишь 8% датасета, состоящего из 100 тысяч данных.

ВЛИЯЕТ ЛИ ВЕС?

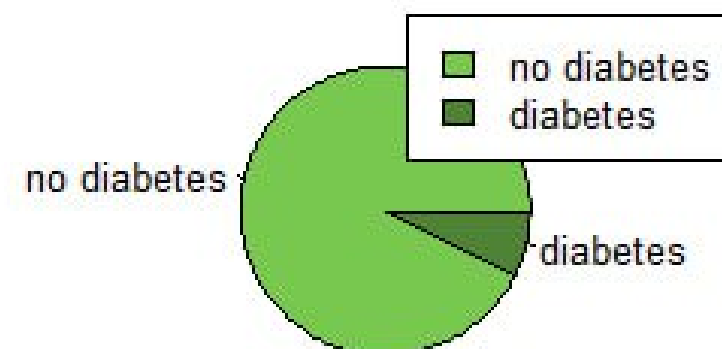
Underweight



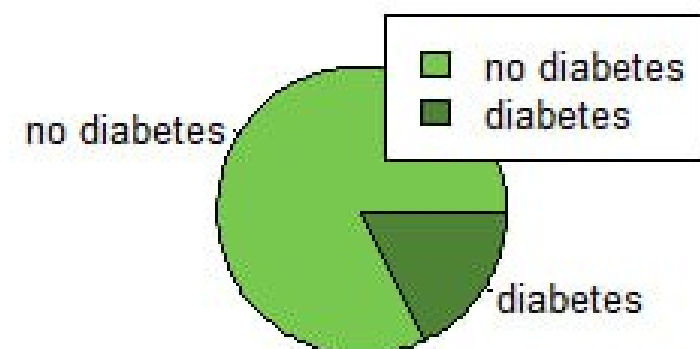
Normal



Overweight



Obesity



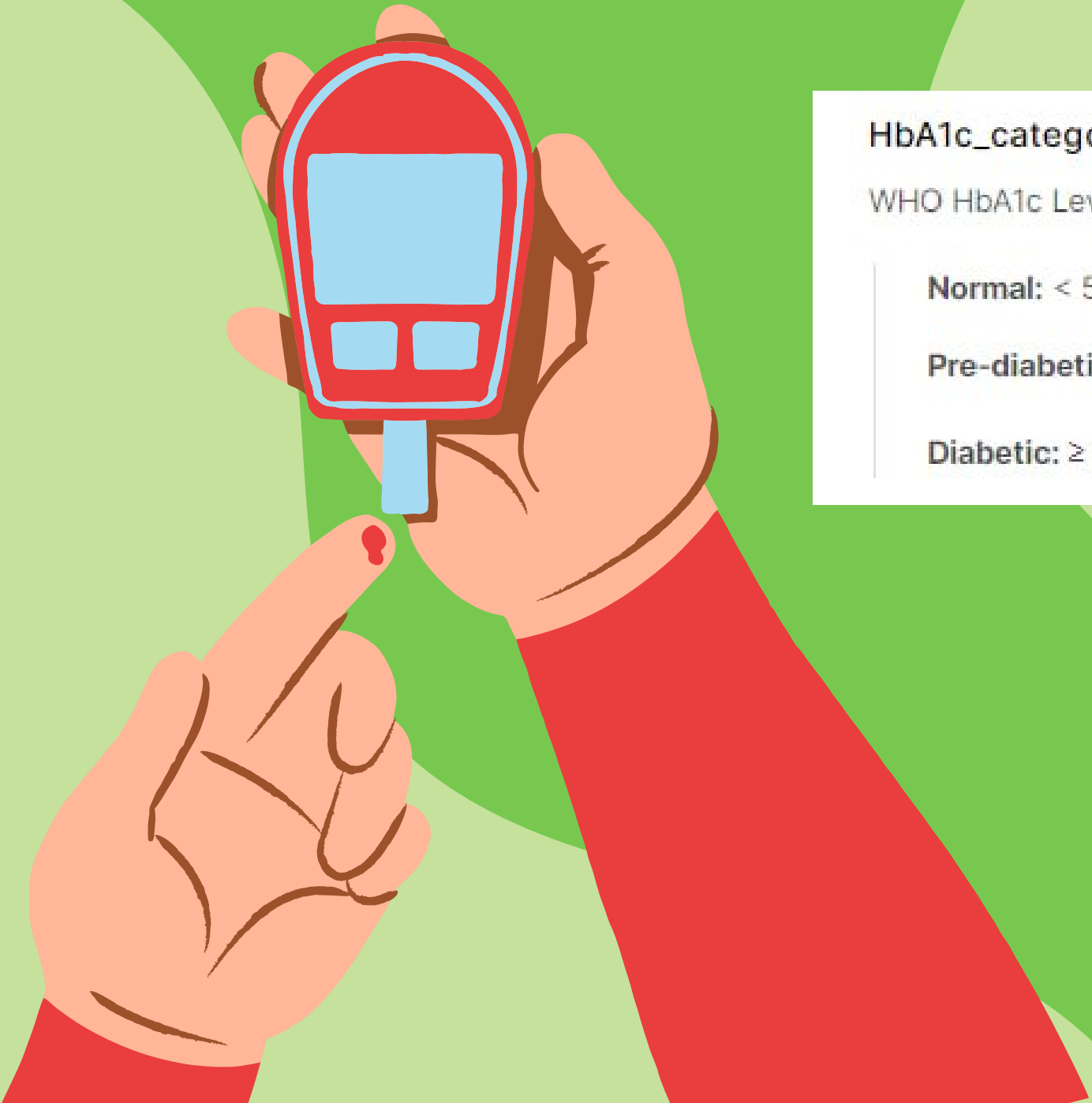
BMI	Category
≤ 18.5	Underweight
18.5 – 24.9	Normal
25 – 29.9	Overweight
≥ 30	Obesity

Пусть лишний вес и не гарантирует вам наличие диабета, однако соблюдение здоровой диеты необходимо!



УРОВЕНЬ САХАРА В КРОВИ

Уровень сахара в крови является важным показателем наличия диабета!



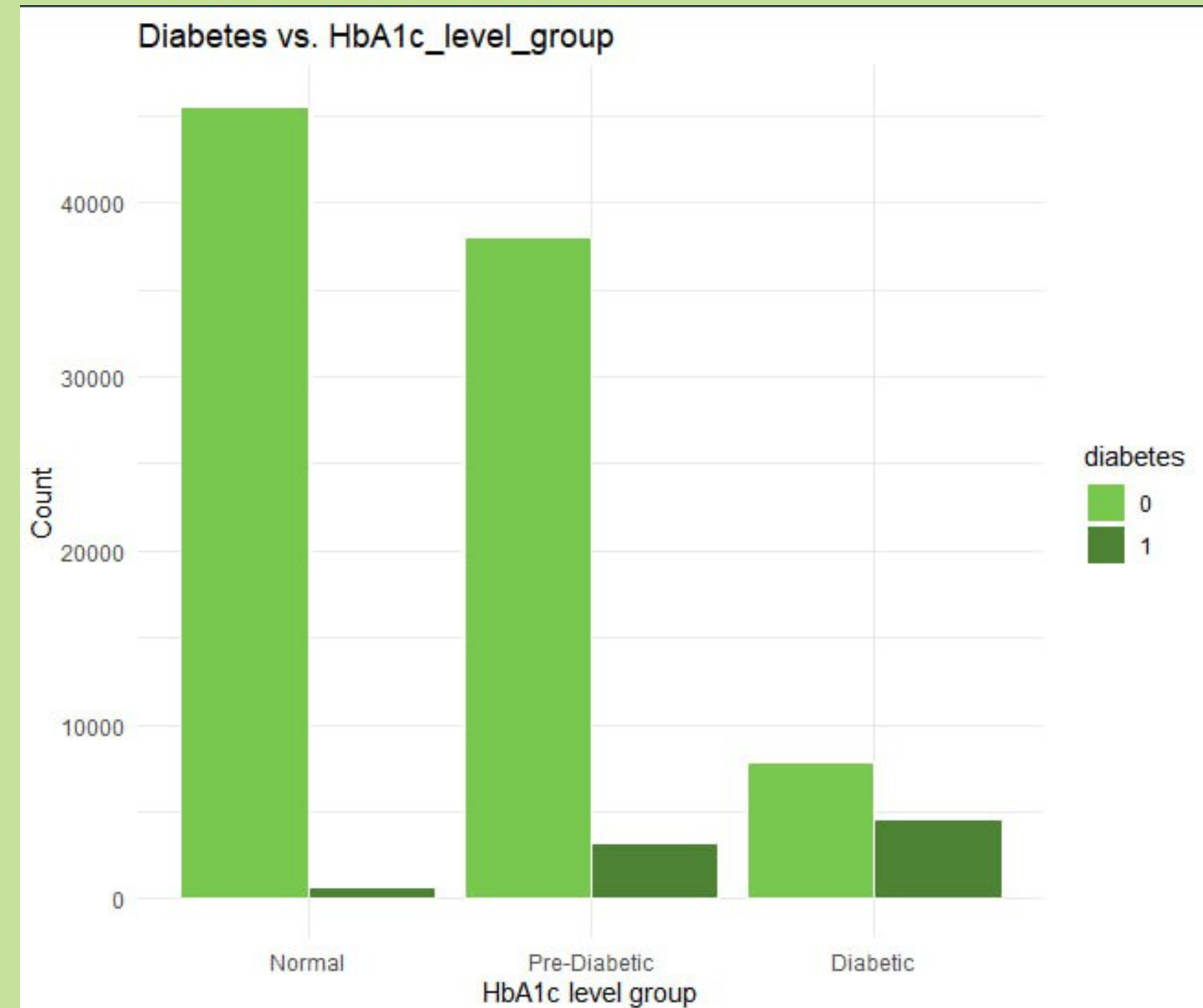
HbA1c_category

WHO HbA1c Levels:

Normal: $< 5.7\%$

Pre-diabetic: $\geq 5.7\%$ and $< 6.5\%$

Diabetic: $\geq 6.5\%$

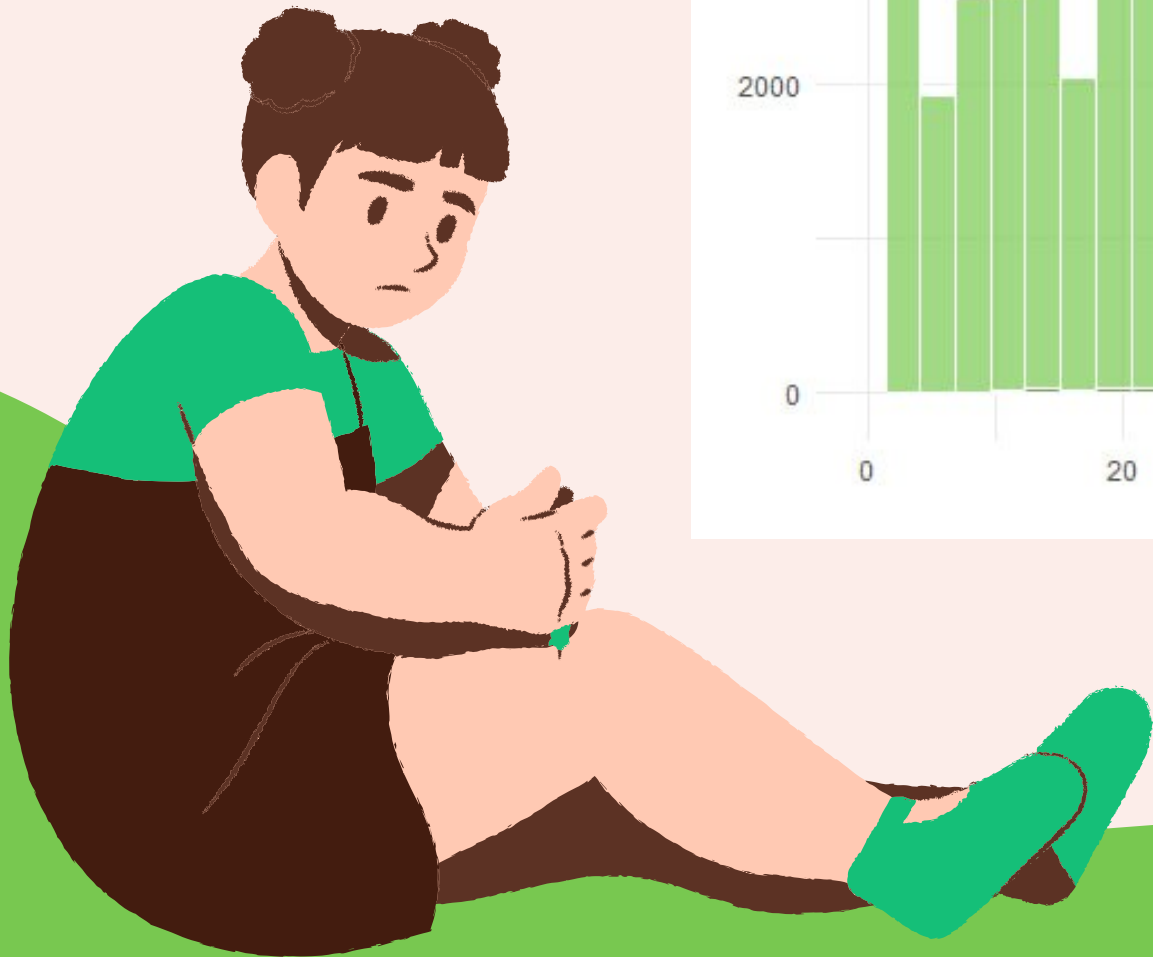
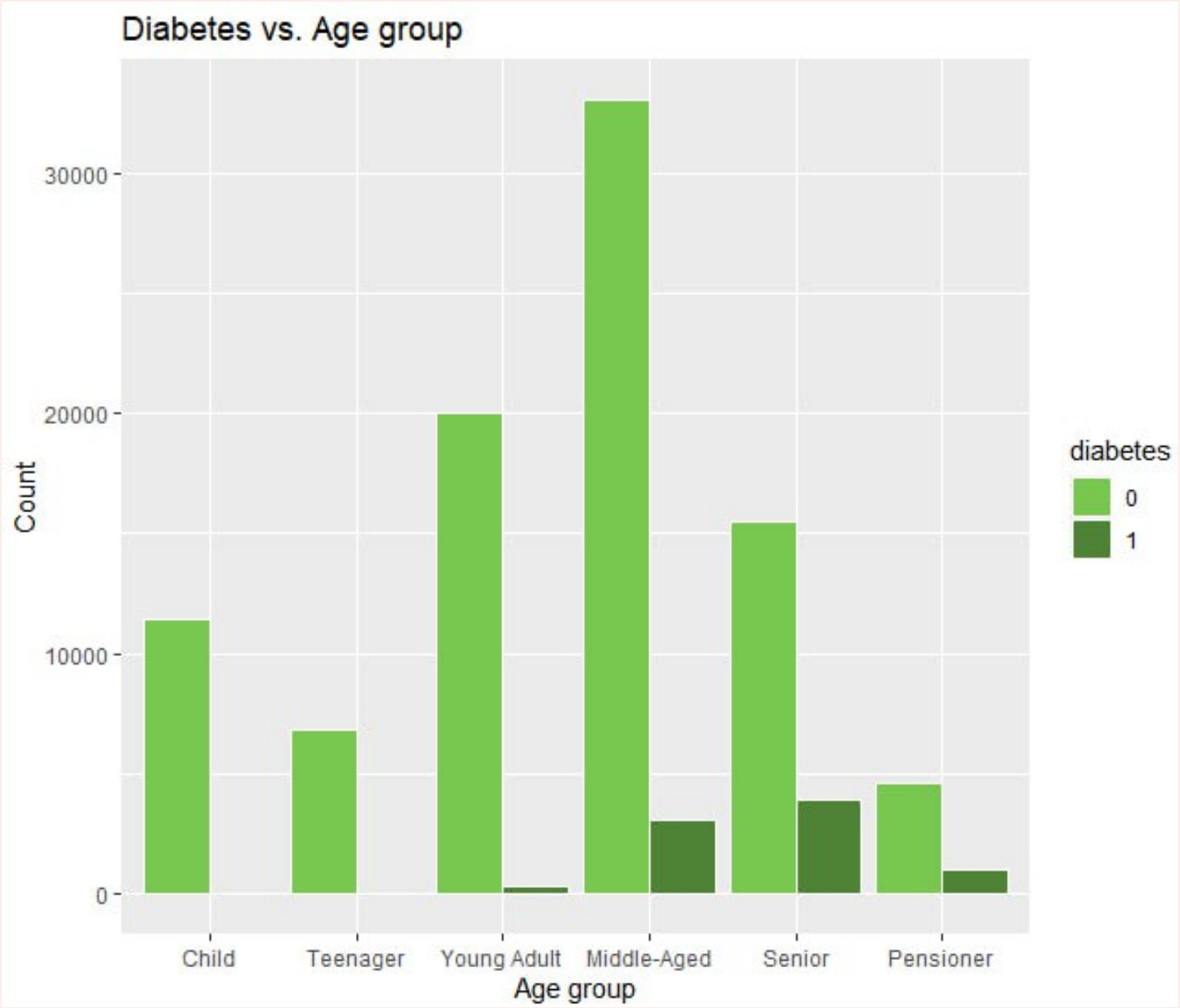
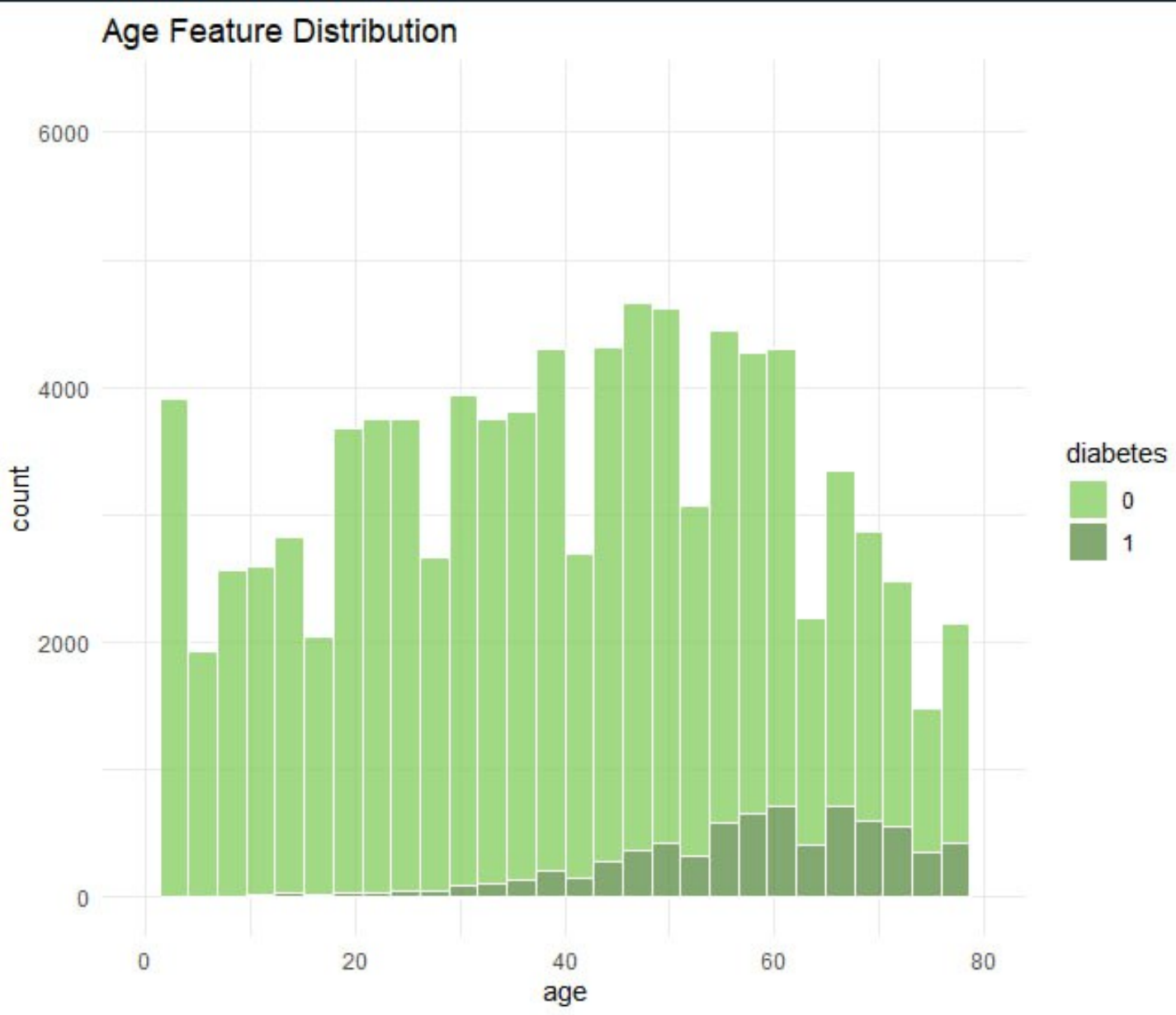


ПОКАЗАТЕЛИ УРОВНЯ ГЛЮКОЗЫ



**Нормальный показатель
глюкозы находится в пределах -
70 -130. Пограничный - 131 - 180.
Высокий 181- 250. А более 250 -
опасный.**

ВЛИЯЕТ ЛИ ВОЗРАСТ?



Confusion Matrix

1	922	1562
0	27266	250
	0	1

**ТОЧНОСТЬ
ЛОГИСТИЧЕСКОЙ
МОДЕЛИ**

Факторы, влияющие на развитие болезни:

1. Возраст (коэффициент 0.0485).
2. Гипертония (коэффициент 0.797).
3. Индекс массы тела (коэффициент 0.0879).
4. Уровень глюкозы (коэффициент 0.0333).
5. Уровень HbA1c (коэффициент 2.36).

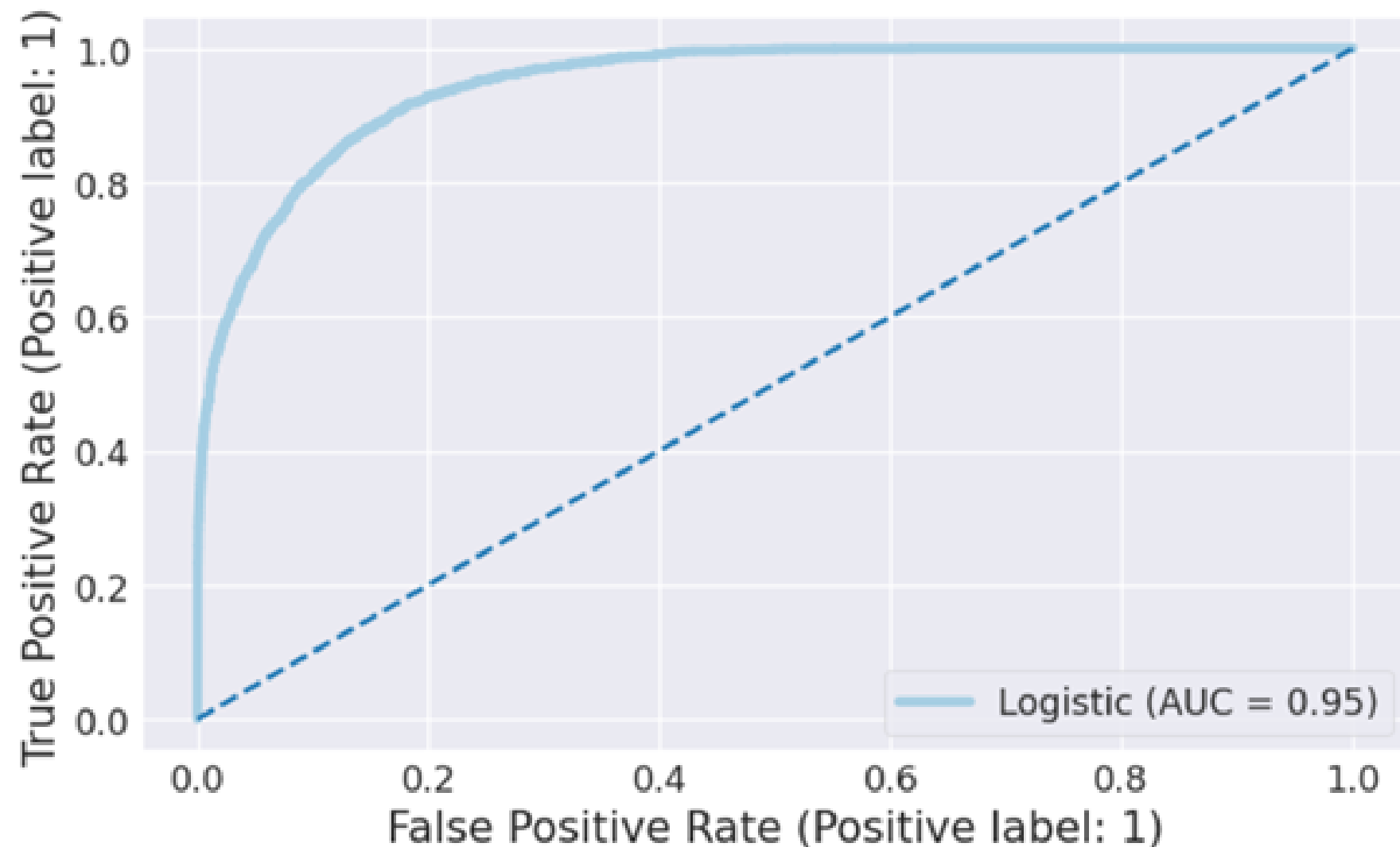
term	estimate	std.error	statistic	p.value
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 (Intercept)	-27.3	0.352	-77.6	0
2 genderMale	0.338	0.0428	7.88	3.16e- 15
3 genderOther	-9.64	113.	-0.0856	9.32e- 1
4 age	0.0485	0.00133	36.5	2.75e-291
5 hypertension	0.797	0.0561	14.2	7.40e- 46
6 bmi	0.0879	0.00306	28.8	6.97e-182
7 smoking_historyever	0.00584	0.109	0.0535	9.57e- 1
8 smoking_historyformer	-0.0393	0.0832	-0.473	6.36e- 1
9 smoking_historynever	-0.173	0.0722	-2.40	1.66e- 2
10 smoking_historyNo Info	-0.713	0.0793	-8.99	2.37e- 19
11 smoking_historynot current	-0.169	0.0987	-1.71	8.70e- 2
12 HbA1c_level	2.36	0.0430	54.9	0
13 blood_glucose_level	0.0333	0.000572	58.2	0

Феноменально!



```
> cat("Accuracy:",  
Accuracy: 0.9603571
```

ТОЧНОСТЬ ЛОГИСТИЧЕСКОЙ МОДЕЛИ



ВЫВОДЫ

Разработанная логистическая модель демонстрирует высокую точность (96%), что свидетельствует о ее эффективности в предсказании диабета на основе медицинских и демографических данных.

Это предоставляет ценный инструмент для раннего выявления рисков, создания персонализированных планов лечения и глубокого анализа факторов, влияющих на вероятность развития диабета.