# Ключевые выводы

1. Выполнено исследование качества данных:

* Выявлены и обработаны скрытые пропуски
* Проведен анализ распределений всех переменных
* Обнаружен дисбаланс классов (65% без диабета и 35% с диабетом)
* Выполнен статистический анализ выбросов методом IQR

1. Проведено исследование взаимосвязей:

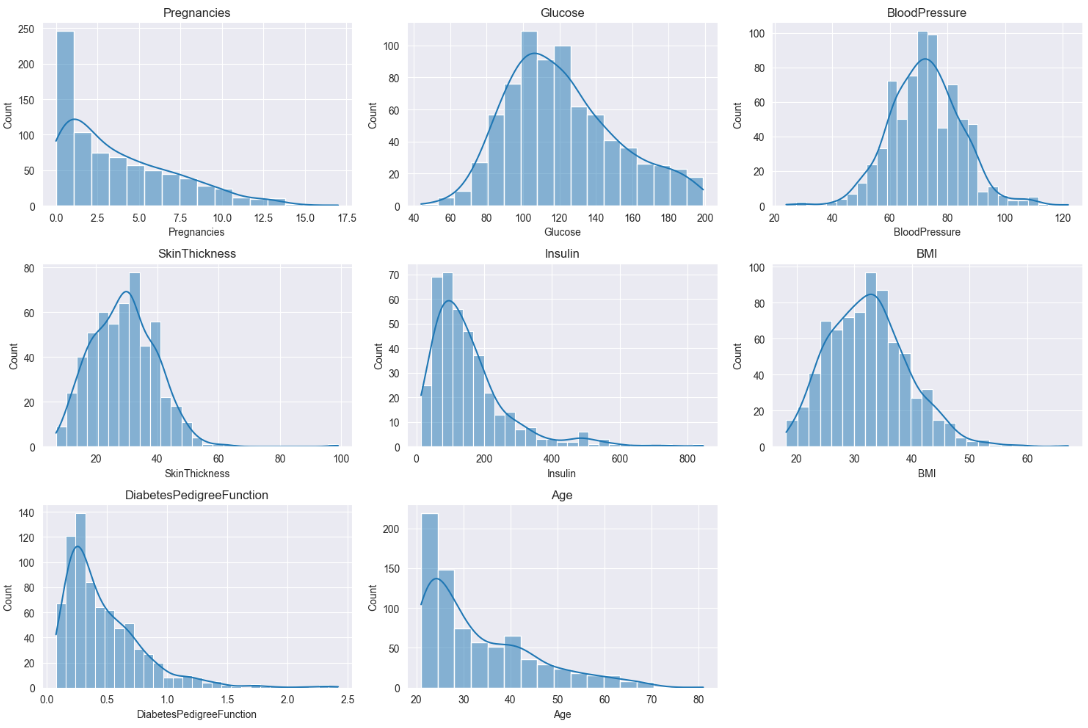
* Применены статистические тесты для анализа значимости признаков
* Построена корреляционная матрица для выявления взаимосвязей (уровень глюкозы наиболее важный предиктор диабета)

1. Осуществлена предобработка данных:

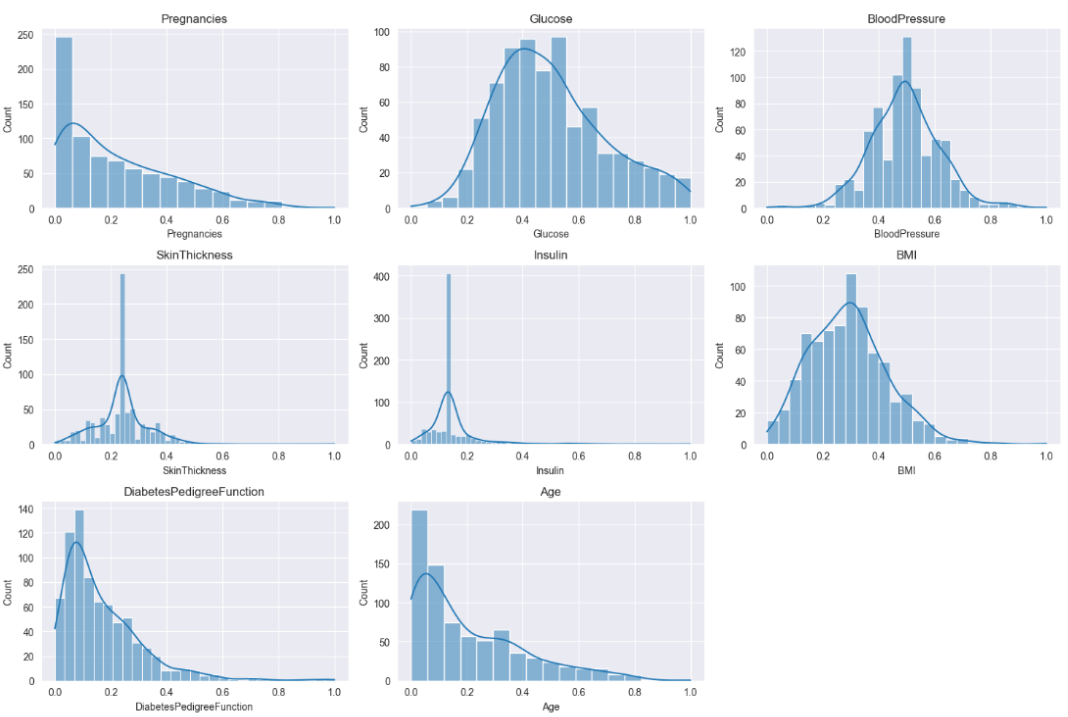
* Импутация пропусков с использованием медианы и среднего
* Масштабирование признаков
* Обработка выбросов методом винсоризации

# Ключевые графики

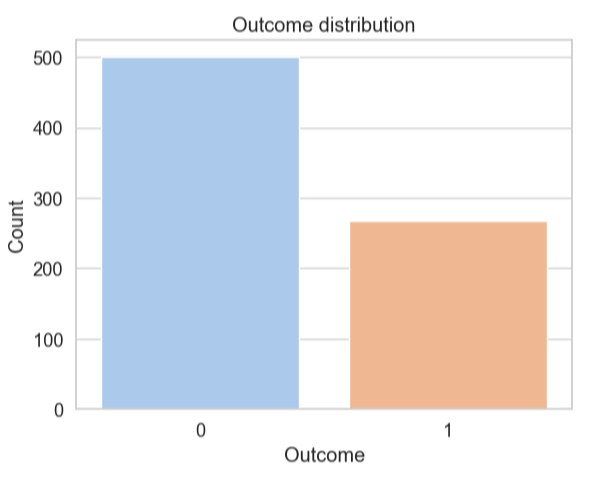
## До нормализации:



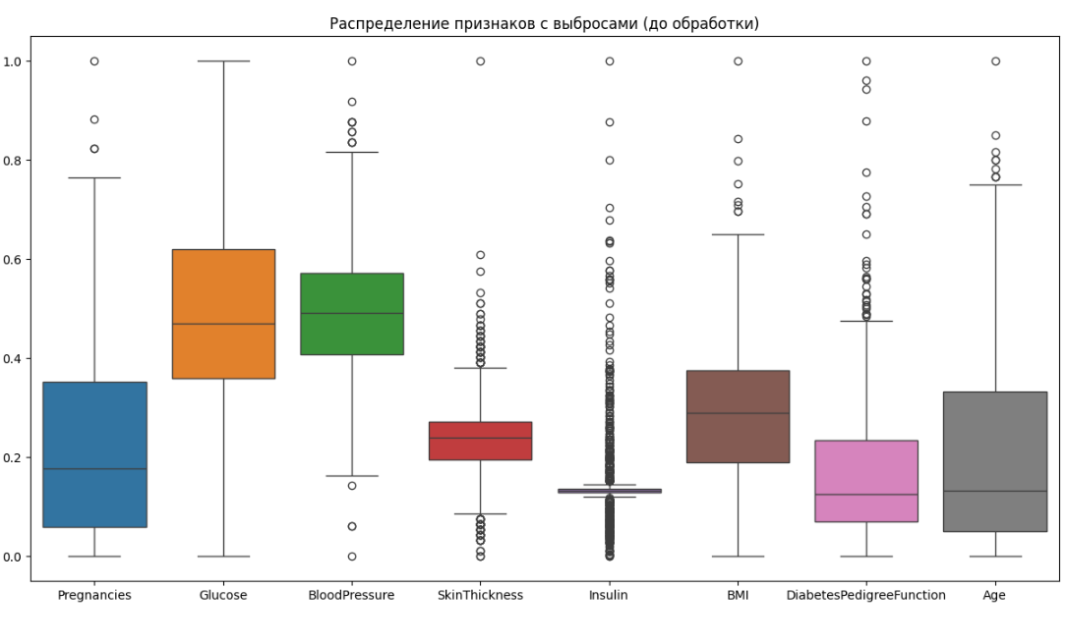
## После нормализации:



## Распределение классов:



## Распределение выбросов:



## Корреляционная матрица:

