

Университет ИТМО

Практическая работа №3
по дисциплине «Визуализация и моделирование»

Автор: Никулина Ксения Григорьевна

Поток: ВИМ 1.1

Группа: К3221

Факультет: ИКТ

Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

1 Heart Failure Prediction

Сердечно-сосудистые заболевания являются причиной смерти номер 1 во всем мире, забирая в среднем 17.9 миллионов жизней в год, что составляет 31% от всех смертей по всему миру. Остановка сердца - результат сердечно-сосудистых заболеваний и данный датасет содержит 12 показателей, которые можно использовать для предсказания смертности от остановки сердца.

Название столбца	Описание данных	Тип данных	Шкала
Возраст	Возраст пациента	Integer	Интерв
Анемия	Снижение гемоглобина	Binary	Номин
Креатинкиназа	Уровень фермента	Integer	Относит
Диабет	Наличие диабета	Binary	Номин
Фракция выброса	% крови при сокращении	Integer	Относит
Высокое кровяное давление	Гипертония	Binary	Номин
Тромбоциты	Тромбоциты в крови	Float	Номин
Креатинин	Уровень креатинина в крови	Float	Номин
Натрий	Уровень натрия в крови	Integer	Относит
Пол	Пол пациента	Binary	Номин
Курение	Курит ли пациент	Binary	Номин
Время	Кол-во дней наблюдения	Integer	Относит
Смерть	Умер ли пациент	Binary	Номин

Начнем работу с обработки дата сета при помощи новых инструментов. Сразу видно, что во всех столбцах отсутствуют пустые значения. Число 13 в крайнем правом столбце означают, что минимальное количество заполненных ячеек в строке - 13, максимальное - 13.

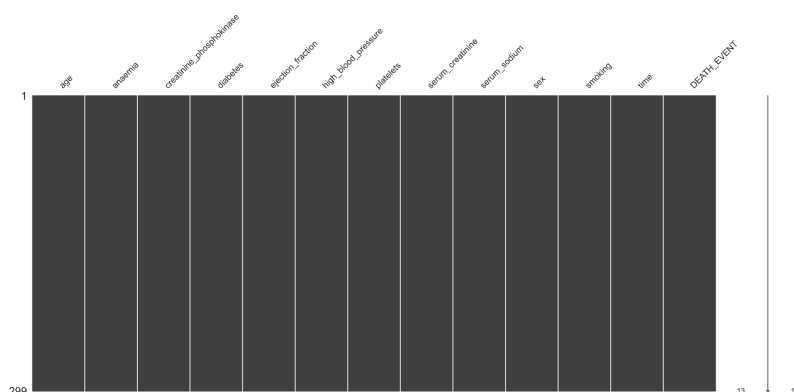


Рис. 1 Результаты обработки

Следующее изображение показывает количество данных в столбцах. Так как пустые значения отсутствуют, то столбцы заполнены одинаково полностью.

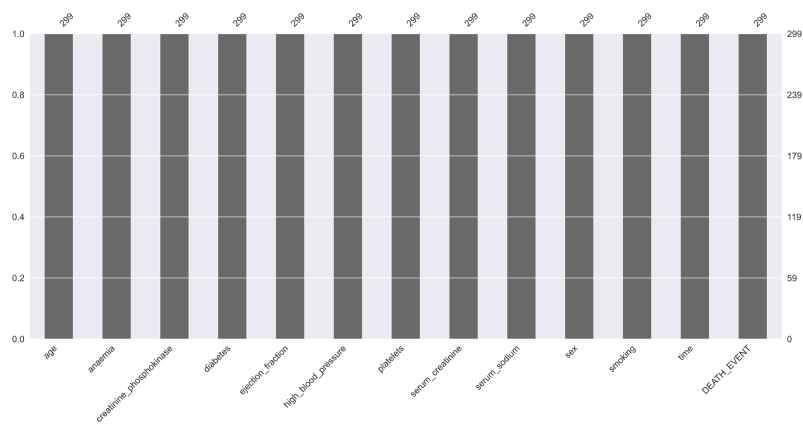


Рис. 2 Результаты обработки

Матрица смежности показывает то, что все столбцы не коррелируют друг от друга.

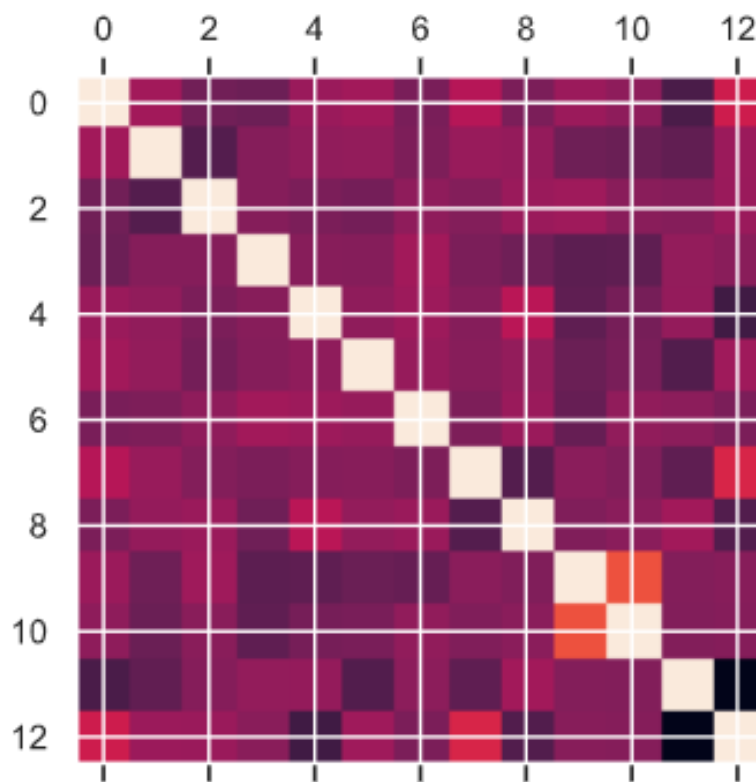


Рис. 3 Результаты обработки

Гипотезы:

1. **Количество пациентов в зависимости от возраста растет линейно.**

Воспользовавшись библиотекой seaborn мы построили распределение количества пациентов по возрасту. Наглядно убедились в том, что график не растет линейно, а выглядит скачкообразно. Гипотеза не подтвердилась.



Рис. 4 Результат первой гипотезы

2. **Чем моложе человек, тем меньше у него шанс умерет от сердечной недостаточности**

По данным визуализации мы видим совершенно обратное: больше всего смертей приходится на средний возраст - 60 лет, а меньше всего на 80+. Гипотеза не подтвердилась.

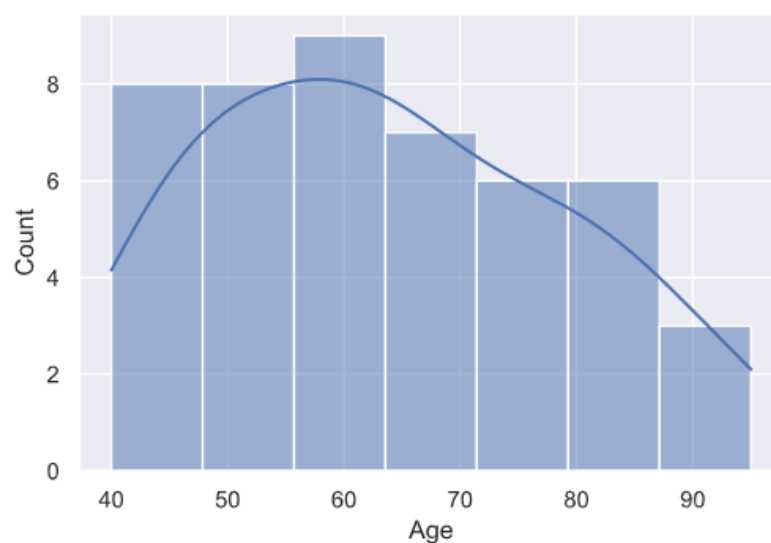


Рис. 5 Результат второй гипотезы

3. **Количество женщин, в более старшем возрасте (60+) под наблюдением превосходит мужчин**

По данным графика можно увидеть, что количество мужчин и женщин либо совпадают, либо количество пациентов мужского пола превосходит. Гипотеза не подтвердилась.

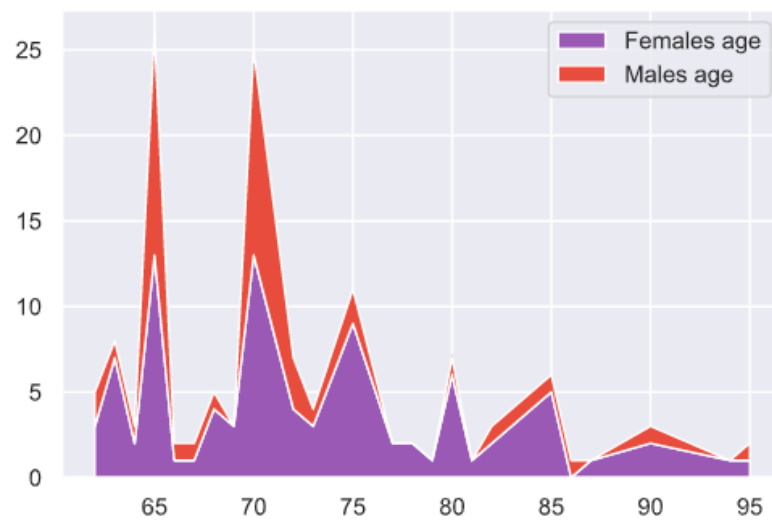


Рис. 6 Результат третьей гипотезы

4. Пациенты с наличием гипертонии умирают чаще от сердечной недостаточности, чем те, у которых ее нет

Согласно круговой диаграмм можно увидеть, что количество умерших людей с гипертонией превышает тех, у кого ее не было. Гипотеза подтвердилась.

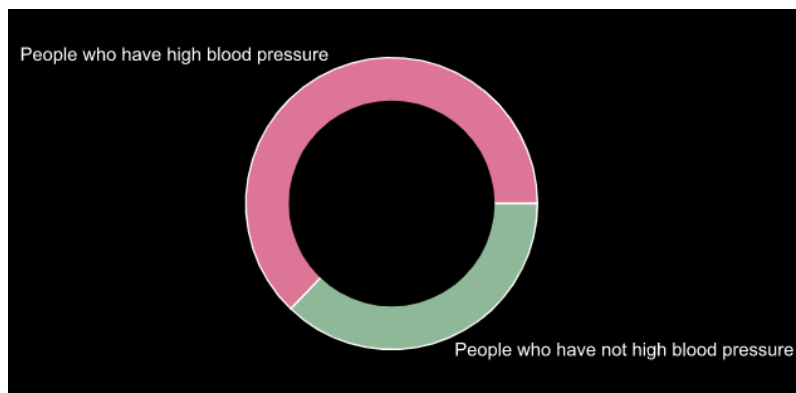


Рис. 7 Результат четвертой гипотезы

5. Наличие курения у людей 55+ лет повышает риск ссб

Согласно визуализированным данным половина курящих пациентов умерла. Поэтому с уверенностью можно сказать, что курение повышает риск сердечно сосудистых заболеваний.

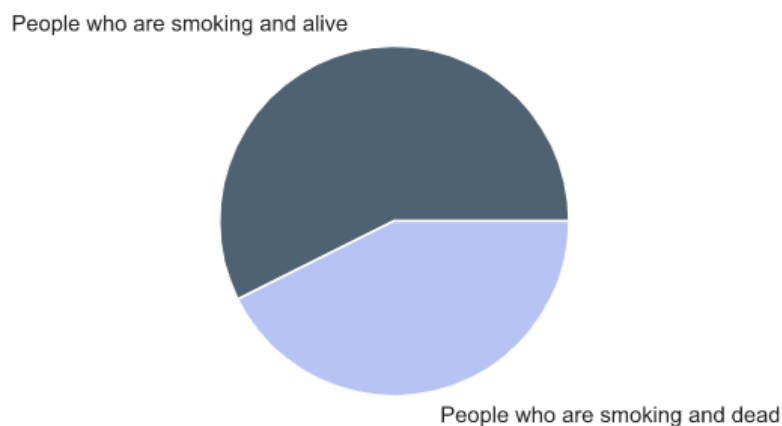


Рис. 8 Результат пятой гипотезы