

# **Asfendiyarov Kazakh National Medical University**

**Тема исследования: Практическое применение  
Автоматизированной системы научных исследований в медицине,  
здравоохранении и смежных областях**

**Проект: Анализ факторов риска сердечно сосудистых заболеваний  
и прогноз исходов лечения при помощи методов Машинного  
Обучения**

**Раздел I. Исследовательский анализ данных (EDA)**

**Автор исследования: Dr. Alexander Wagner (Berlin)**





## Содержание

Предисловие .....	5
Материалы и методы .....	5
Этапы исследования.....	5
Исследовательский анализ данных (EDA) .....	6
Введение в EDA.....	6
Набор данных для анализа .....	9
Преобразование категориальной переменной в числовую .....	11
Результаты EDA .....	13
Ниже приведены результаты проведения ЕДА в форме таблиц и графиков.....	13
Таблица №1. Исходные данные (выборка) для кардио-пациентов .....	13
Таблица №2. Описательная статистика читловых переменных набора данных для всех кардио-пациентов.....	14
График №1. BoxPlot для всех числовых переменных по классу Заболевание (HeartDisease).....	14
Таблица №3 .....	15
График №2. BoxPlot для всех числовых переменных .....	17
Таблица №4.....	17
График №3. Распределение числовых переменных по классу пол (Sex) .....	19
Таблица №5.....	20
График №4. Гистограммы распределения для всех числовых переменных по классу Заболевание (HeartDisease).....	20
Таблица №6.....	21
График №5. Гистограммы распределения для всех числовых переменных по классу пол (Sex) .....	23
Таблица №7 .....	23
График №6. Двумерное распределение переменной 'HeartDisease' по классам Sex и RestingBP.....	25
Таблица №8 .....	25
График №7. График распределение переменной Cholesterol по Возрасту (Age).....	27
Таблица №9 .....	27
График №8. Гистограммы распределения для всех числовых переменных, представленные на одном графике.....	29
Таблица №10.....	29
График №9. Матрица корреляции Пирсона для числовых переменных .....	31
Таблица №11 .....	31
График №10. Гистограммы распределения для всех числовых переменных, в виде субграфиков на одной панели.....	33
Таблица №12 .....	34
График №11. Распределение переменной Cholesterol по классу (Age, Sex, FastingBS) в виде 4 субграфиков на одной панели .....	39



Таблица №13 .....	40
График №12. Распределение по возрасту (Age) для переменных: 'Sex', 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST_Slope', 'HeartDisease' в форме Виалин-графиков .....	43
Таблица №14 .....	44
График №13. Распределение всех числовых переменных в виде субграфиков по классу 'HeartDisease' на одной панели .....	48
График №14. Распределение всех числовых переменных в виде столбиковых диаграмм как субграфиков по классу 'HeartDisease' на одной панели .....	49
График №15. Распределение численности пациентов по всем переменным в виде столбиковых диаграмм как субграфиков по классу пол (Sex) на одной панели .....	50
График №16. Биполярное распределение переменной Cholesterol по возрасту 'Age' .....	51
График №17. Биполярное распределение разного графического типа всех численных переменных на одной панели .....	53
График №18. Плотность распределения переменной Cholesterol по классу пол (Sex) .....	55
График №19. Распределение переменной Age возрастным группам в виде столбиковых диаграм по классу 'HeartDisease' .....	57
График №20. Распределение переменной Cholesterol по возрастным группам в виде столбиковых диаграм по классу 'HeartDisease' .....	59
График №21. Распределение переменной RestingBP по возрастным группам в виде столбиковых диаграм по классу 'HeartDisease' .....	61
График №22. Распределение переменной MaxHR по возрастным группам в виде столбиковых диаграм по классу 'HeartDisease' .....	63
График №23. Распределение переменной Oldpeak по возрастным группам в виде столбиковых диаграм по классу 'HeartDisease' .....	65
График №24. Торт-диаграмма распределения пациентов по класс-переменной HeartDisease .....	67
График №25. Торт-диаграмма распределения пациентов по класс-переменной Sex .....	69
График №26. Столбиковая диаграмма распределения переменных: 'Sex', 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST_Slope', 'HeartDisease' по категориям .....	71
График №27. Блок-бокс диаграмма распределения переменных: 'Sex', 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST_Slope', 'HeartDisease' в виде субграфиков на одной панели по категориям .....	73
График №28. Столбиковая диаграмма распределения переменных: 'Sex', 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST_Slope', 'HeartDisease' по категориям .....	75
График №29. Секторная диаграмма распределения переменной ChestPainType .....	77
График №30. Секторная диаграмма (2-го типа) распределения переменной ST_Slope ..	79
График №31. Комбинированная диаграмма распределения переменной пол (Sex) .....	81
График №32. Комбинированная диаграмма распределения переменной HeartDisease ....	83
График №33. Столбиковая диаграмма распределения переменных: 'Sex', 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST_Slope' по категориям и классу 'HeartDisease' .....	85



График №34. Столбиковая диаграмма распределения переменных: 'ChestPainType','FastingBS','RestingECG','ExerciseAngina', 'ST_Slope' по категориям и классу 'Sex' .....	87
График №35. Столбиковая диаграмма распределения переменных: 'Sex','FastingBS','RestingECG','ExerciseAngina','ST_Slope', 'HeartDisease' по категориям и классу 'ChestPainType' .....	89
График №36. Комбинированная (столбиковая и секторная) диаграмма распределения переменной пол (Sex).....	91
График №37. Комбинированная (столбиковая и секторная) диаграмма распределения переменной ChestPainType .....	93
График №38. Комбинированная (столбиковая и секторная) диаграмма распределения переменной RestingECG .....	95
График №39. Комбинированная (столбиковая и секторная) диаграмма распределения переменной ExerciseAngina.....	97
График №40. Комбинированная (столбиковая и секторная) диаграмма распределения переменной ST_Slope.....	99
График №41. Комбинированная (столбиковая и секторная) диаграмма распределения переменной Cholesterol_Category .....	101
График №42. Комбинированная (столбиковая и секторная) диаграмма распределения переменной RestingBP_Category .....	103
График №43. Двойная секторная диаграмма (Sunburst, 6 субграфиков) распределения пар переменных: ['ChestPainType', 'FastingBS'], ['ST_Slope', 'RestingECG'], ['ExerciseAngina', 'ChestPainType'] по классу пол (Sex) .....	105
График №44. Столбиковая диаграмма распределения переменных: RestingECG, ChestPainType (2 субграфика на одной панели) по классу пол (Sex).....	107
График №45. Столбиковая диаграмма распределения переменных: ST_Slope, ExerciseAngina (2 субграфика на одной панели) по классу пол (Sex).....	109
Заключение по результатам EDA .....	110



## Предисловие

Данная работа посвящена проблеме проведения научного исследования и создания научного отчета в медицине и здравоохранении при помощи Автоматизированной Системы Научных Исследований. Краткая характеристика системы приведена в данном разделе. Описание научного исследования, выполненное по установленным международным нормам, правилам и рекомендациям, приводится в последующих разделах данного документа.

## Материалы и методы

### Этапы исследования

Наше исследование состоит из двух этапов.

#### Этап первый. Исследовательский анализ данных (EDA)

Исследовательский анализ включает в себя изучение данных и поиск связей между переменными, которые ранее были неизвестны. Вот что вам нужно знать:

- EDA помогает обнаруживать взаимосвязи между показателями в данных, которые не являются доказательством существования корреляции, как обозначается фразой «Корреляция не подразумевает причинно-следственную связь».
- Это полезно для обнаружения новых связей и формирования гипотез. Он управляет планированием проектирования и сбором данных.

Для предварительной обработки набора данных мы классифицировали все атрибутивные признаки по различным категориям, например, категориальные признаки, непрерывные признаки и дискретные признаки. Мы определили различные правила предварительной обработки для этих различных типов атрибутов. Для категориальных переменных мы применили кодирование меток [32], а также одно горячее кодирование [26] для предварительной обработки этих переменных. Для непрерывных атрибутов мы классифицировали набор данных по диапазонам, а затем применили кодировку меток для этих определенных подклассов. Для некоторых категориальных и непрерывных переменных, содержащих несколько значений, мы применили одно горячее кодирование, чтобы упростить управление значениями этих атрибутов. Одно горячее кодирование является одним из лучших решений для управления несколькими значениями и предварительной обработки тех атрибутов, которые содержат более одного параметра. В нашем наборе данных также есть атрибуты с двоичным значением. Для таких атрибутов, содержащих ровно два значения, мы преобразовали их в двоичную форму (0 и 1), обозначив 0 как No, 1 как Yes.

#### Этап второй. Моделирование и Анализ

Моделирование и Анализ заключается в проведение при помощи методов и алгоритмов (МЛ) исследования влияния факторов риска на заболеваемость пациентов ССЗ.

Ниже в последующих разделах приводятся детальное изложение обоих этапов исследования.



## Исследовательский анализ данных (EDA)

### Введение в EDA

Исследовательский анализ данных (*Exploratory data analysis, EDA*) используется для анализа и исследования наборов данных и обобщения их основных характеристик.

EDA помогает определить, как лучше всего манипулировать входными данными для получения необходимых ответов, упрощая специалистам по обработке и анализу данных обнаружение закономерностей, выявление аномалий, проверку гипотез или предположений.

EDA в основном используется для того, чтобы увидеть, что данные могут выявить за пределами формального моделирования или проверки гипотез, и обеспечивает лучшее понимание переменных набора данных и отношений между ними. Это также может помочь определить, подходят ли статистические методы, которые вы рассматриваете для анализа данных.

Основная цель EDA — помочь взглянуть на данные, прежде чем делать какие-либо предположения. Это может помочь выявить очевидные ошибки, а также лучше понять закономерности в данных, обнаружить выбросы или аномальные события, найти интересные связи между переменными.

Специалисты по обработке и анализу данных могут использовать исследовательский анализ, чтобы убедиться, что результаты, которые они получают, являются достоверными и применимыми к любым желаемым бизнес-результатам и целям. EDA также помогает заинтересованным сторонам, подтверждая, что они задают правильные вопросы. EDA может помочь ответить на вопросы о стандартных отклонениях, категориальных переменных и доверительных интервалах.

Конкретные статистические функции и методы, которые можно выполнять с помощью инструментов EDA, включают:

- Методы кластеризации и уменьшения размерности, которые помогают создавать графические представления многомерных данных, содержащих множество переменных.
- Одномерная визуализация каждого поля в необработанном наборе данных со сводной статистикой.
- Двумерные визуализации и сводная статистика, которые позволяют оценить связь между каждой переменной в наборе данных и целевой переменной, которую вы просматриваете.
- Многомерные визуализации для сопоставления и понимания взаимодействий между различными полями данных.
- Кластеризация K-средних — это метод кластеризации в , при котором точки данных распределяются по K-группам, т.е. по количеству кластеров, в зависимости от расстояния от центроида каждой группы. Точки данных, ближайшие к определенному центроиду, будут объединены в одну категорию. Кластеризация K-средних обычно используется для сегментации рынка, распознавания образов и сжатия изображений.
- Прогностические модели, такие как линейная регрессия, используют статистику и данные для прогнозирования результатов.



Существует четыре основных типа EDA:

**Одномерный неграфический.** Это простейшая форма анализа данных, при которой анализируемые данные состоят всего из одной переменной. Поскольку это одна переменная, она не имеет отношения к причинам или отношениям. Основной целью одномерного анализа является описание данных и поиск закономерностей, которые в них существуют.

**Одномерная графика.** Неграфические методы не дают полной картины данных. Поэтому требуются графические методы. К распространенным типам одномерной графики относятся:

Штамбовые и листовые графики, которые показывают все значения данных и форму распределения.

Гистограммы, гистограмма, на которой каждый столбец представляет частоту (количество) или долю (количество/общее количество) вариантов для диапазона значений.

Ящичковые диаграммы, которые графически изображают сводку из пяти чисел минимума, первого квартиля, медианы, третьего квартиля и максимума.

**Многомерные неграфические:** Многомерные данные возникают из нескольких переменных. Многомерные неграфические методы САПР обычно показывают взаимосвязь между двумя или более переменными данных с помощью перекрестной таблицы или статистики.

**Многовариантная графика:** Многомерные данные используют графику для отображения связей между двумя или более наборами данных. Наиболее часто используемым рисунком является сгруппированная линейчатая диаграмма или линейчатая диаграмма, где каждая группа представляет один уровень одной из переменных, а каждая полоса в группе представляет уровень другой переменной.

К другим распространенным типам многомерной графики относятся:

Точечная диаграмма, которая используется для отображения точек данных по горизонтальной и вертикальной оси, чтобы показать, насколько одна переменная подвержена влиянию другой.

Многомерная диаграмма, представляющая собой графическое представление взаимосвязей между факторами и реакцией.

Запустите диаграмму, которая представляет собой линейный график данных, построенный во времени.

Пузырьковая диаграмма, представляющая собой визуализацию данных, отображающую несколько кругов (пузырьков) на двумерном графике.

Тепловая карта, представляющая собой графическое представление данных, где значения изображены цветом.





## Набор данных для анализа

Краткое описание переменных набора входных данных[95]

**Age:** age of the patient [years]

**Sex:** sex of the patient [M: Male, F: Female]

**ChestPainType:** chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]

**RestingBP:** resting blood pressure [mm Hg]

**Cholesterol:** serum cholesterol [mm/dl]

**FastingBS:** fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]

**RestingECG:** resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]

**MaxHR:** maximum heart rate achieved [Numeric value between 60 and 202]

**ExerciseAngina:** exercise-induced angina [Y: Yes, N: No]

**Oldpeak:** oldpeak = ST [Numeric value measured in depression]

**ST\_Slope:** the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]

**HeartDisease:** output class [1: heart disease, 0: Normal]

Наш набор данных содержит в общей сложности 6 числовых переменных:

**Age, RestingBP, Cholesterol, MaxHR, Oldpeak, HeartDisease**

В нашем наборе данных имеется также 6 категориальных переменных:

**Sex, ChestPainType, FastingBS, RestingECG, ExerciseAngina, ST\_Slope**





## Преобразование категориальной переменной в числовую

Категориальные значения — это тип данных, которые могут быть сгруппированы в различные категории, такие как пол, цвет или жанр. Числовые значения — это данные, которые имеют числовое значение, например возраст, рост или цена. Нам необходимо преобразовать категориальные значения в числовые в САПР по нескольким причинам:

Требования к моделированию: При планировании создания прогнозных моделей или моделей машинного обучения (МО), необходимы числовые представления. Многие алгоритмы машинного обучения, особенно основанные на математических уравнениях, работают с числовыми входными данными.

Визуализация: Числовые данные часто легче визуализировать и интерпретировать. Графики, диаграммы и другие визуализации обычно используются для получения аналитических сведений о данных, а числовые представления облегчают создание осмысленных визуализаций данных.

Статистический анализ: Некоторые статистические тесты и анализы предполагают наличие числовых данных. Например, коэффициенты корреляции, регрессионный анализ и другие статистические методы предназначены для работы с числовыми переменными.

Согласованность типов данных: преобразование категориальных переменных в числовые представления помогает поддерживать согласованность типов данных в наборе данных. Такая согласованность может упростить процесс анализа данных и сделать его более понятным.

Совместимость с методами анализа: Многие статистические методы и методы машинного обучения требуют числовых входных данных. Преобразование категориальных переменных в числовые представления позволяет применять к набору данных более широкий спектр методов анализа.

Существует два распространенных метода преобразования категориальных переменных в числовые:

One-Hot Encoding: Этот метод создает двоичные столбцы для каждой категории, представляющие наличие или отсутствие этой категории.

Кодировка метки: Этот метод присваивает каждой категории уникальное целое число.

Однако горячее кодирование более распространено для двоичных категориальных переменных, так как оно явно представляет каждую категорию независимо.

Здесь мы используем метод One-Hot Encoding для преобразования данных.





## Результаты ЕДА

Ниже приведены результаты проведения ЕДА в форме таблиц и графиков.

**Таблица №1. Исходные данные (выборка) для кардио-пациентов**

Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
M	ATA	140	289	0	Normal	172	N	0	Up	0
F	NAP	160	180	0	Normal	156	N	1	Flat	1
M	ATA	130	283	0	ST	98	N	0	Up	0
F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
M	NAP	150	195	0	Normal	122	N	0	Up	0
M	NAP	120	339	0	Normal	170	N	0	Up	0
F	ATA	130	237	0	Normal	170	N	0	Up	0
M	ATA	110	208	0	Normal	142	N	0	Up	0
M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
F	ATA	120	284	0	Normal	120	N	0	Up	0
F	NAP	130	211	0	Normal	142	N	0	Up	0
M	ATA	136	164	0	ST	99	Y	2	Flat	1
M	ATA	120	204	0	Normal	145	N	0	Up	0
M	ASY	140	234	0	Normal	140	Y	1	Flat	1
F	NAP	115	211	0	ST	137	N	0	Up	0
F	ATA	120	273	0	Normal	150	N	1.5	Flat	0
M	ASY	110	196	0	Normal	166	N	0	Flat	1
F	ATA	120	201	0	Normal	165	N	0	Up	0
M	ASY	100	248	0	Normal	125	N	1	Flat	1
M	ATA	120	267	0	Normal	160	N	3	Flat	1
F	TA	100	223	0	Normal	142	N	0	Up	0
M	ATA	120	184	0	Normal	142	N	1	Flat	0
F	ATA	124	201	0	Normal	164	N	0	Up	0
M	ATA	150	288	0	Normal	150	Y	3	Flat	1
M	NAP	130	215	0	Normal	138	N	0	Up	0
M	NAP	130	209	0	Normal	178	N	0	Up	0
M	ASY	124	260	0	ST	112	Y	3	Flat	0
M	ATA	120	284	0	Normal	118	N	0	Up	0
F	ATA	113	468	0	Normal	127	N	0	Up	0



**Таблица №2. Описательная статистика числовых переменных набора данных для всех кардио- пациентов**

Variable	count	mean	std	min	25%	50%	75%	max
Age	918	53.51	9.43	28	47	54	60	77
RestingBP	918	132.4	18.51	0	120	130	140	200
Cholesterol	918	198.8	109.38	0	173.25	223	267	603
FastingBS	918	0.23	0.42	0	0	0	0	1
MaxHR	918	136.81	25.46	60	120	138	156	202
Oldpeak	918	0.89	1.07	-2.6	0	0.6	1.5	6.2
HeartDisease	918	0.55	0.5	0	0	1	1	1

**График №1. BoxPlot для всех числовых переменных по классу Заболевание (HeartDisease)**

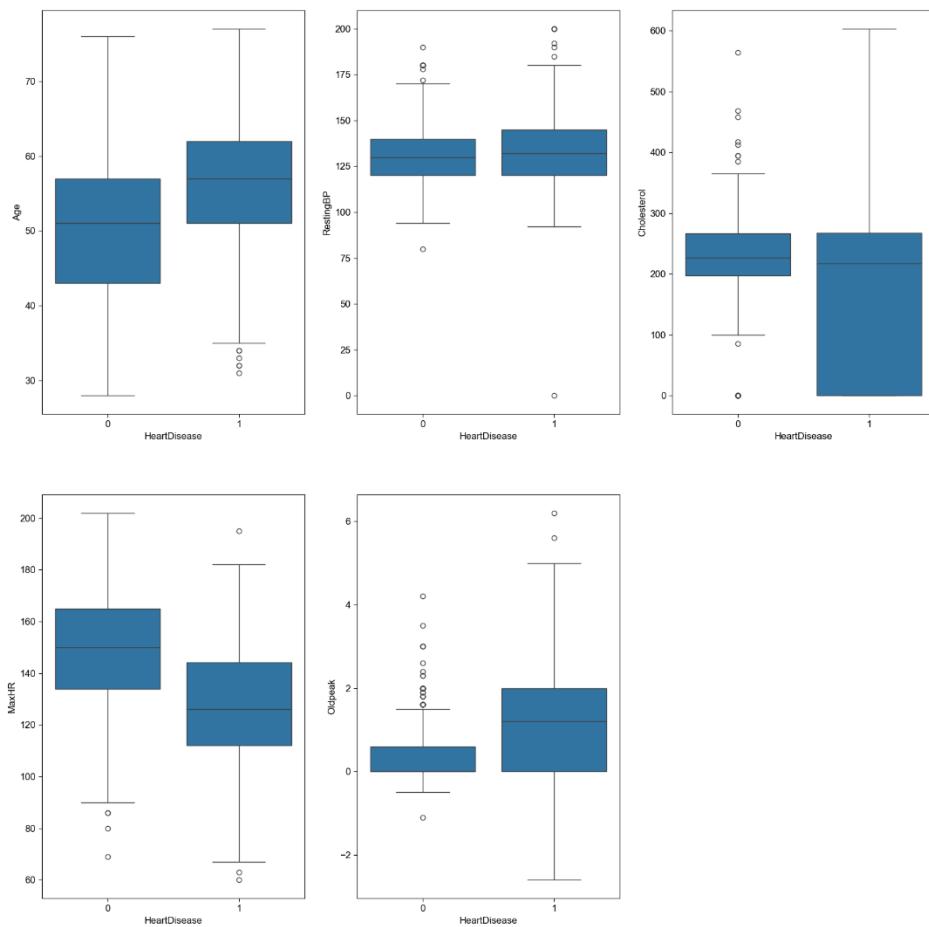




Таблица №3

Analysis: Cholesterol by Geschlecht, Alter								
Female	0 - 29 years	4	132	263	131	211	224	58
	30 - <39 years	49	117	529	412	243	240	68
	40 - <65 years	448	85	603	518	241	233	57
	65 - <75 years	57	110	384	274	243	245	51
	75+ years	6	113	310	197	221	214	77
Male	30 - <39 years	17	160	275	115	207	210	34
	40 - <65 years	148	141	468	327	259	253	58
	65 - <75 years	16	149	564	415	281	267	97
	75+ years	1	197	197	0	197	197	
Total	0 - 29 years	4	132	263	131	211	224	58
	30 - <39 years	66	117	529	412	234	220	63
	40 - <65 years	596	85	603	518	246	238	58
	65 - <75 years	73	110	564	454	251	246	66
	75+ years	7	113	310	197	218	203	70

© Dr. Alexander Wagner. Все права охраняются законом





## График №2. BoxPlot для всех числовых переменных Boxplots for each variable

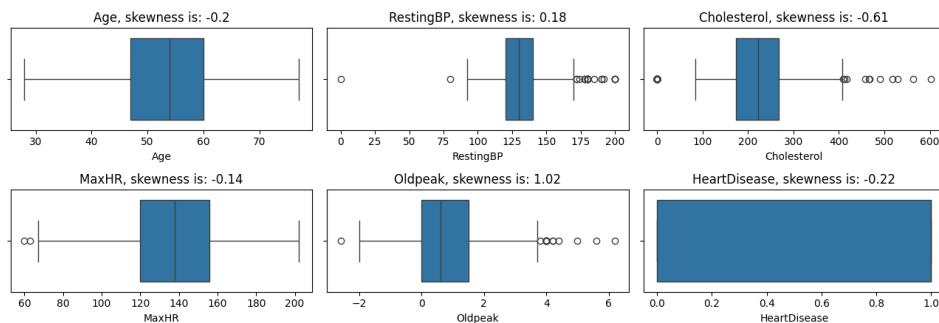


Таблица №4

Analysis: RestingBP by Geschlecht, Alter								
		N	Min	Max	Range	Mean	Median	Std
<b>Female</b>	<b>0 - 29 years</b>	4	120	140	20	130	130	8
	<b>30 - &lt;39 years</b>	57	92	190	98	124	120	17
	<b>40 - &lt;65 years</b>	579	80	200	120	132	130	18
	<b>65 - &lt;75 years</b>	78	100	180	80	140	140	16
	<b>75+ years</b>	6	104	170	66	137	131	25
<b>Male</b>	<b>30 - &lt;39 years</b>	19	94	170	76	121	120	18
	<b>40 - &lt;65 years</b>	155	95	200	105	133	130	19
	<b>65 - &lt;75 years</b>	18	106	178	72	141	148	21
	<b>75+ years</b>	1	140	140	0	140	140	
<b>Total</b>	<b>0 - 29 years</b>	4	120	140	20	130	130	8
	<b>30 - &lt;39 years</b>	76	92	190	98	124	120	17
	<b>40 - &lt;65 years</b>	734	80	200	120	132	130	18
	<b>65 - &lt;75 years</b>	96	100	180	80	140	140	17
	<b>75+ years</b>	7	104	170	66	137	136	22

© Dr. Alexander Wagner. Все права охраняются законом





### График №3. Распределение числовых переменных по классу пол (Sex)

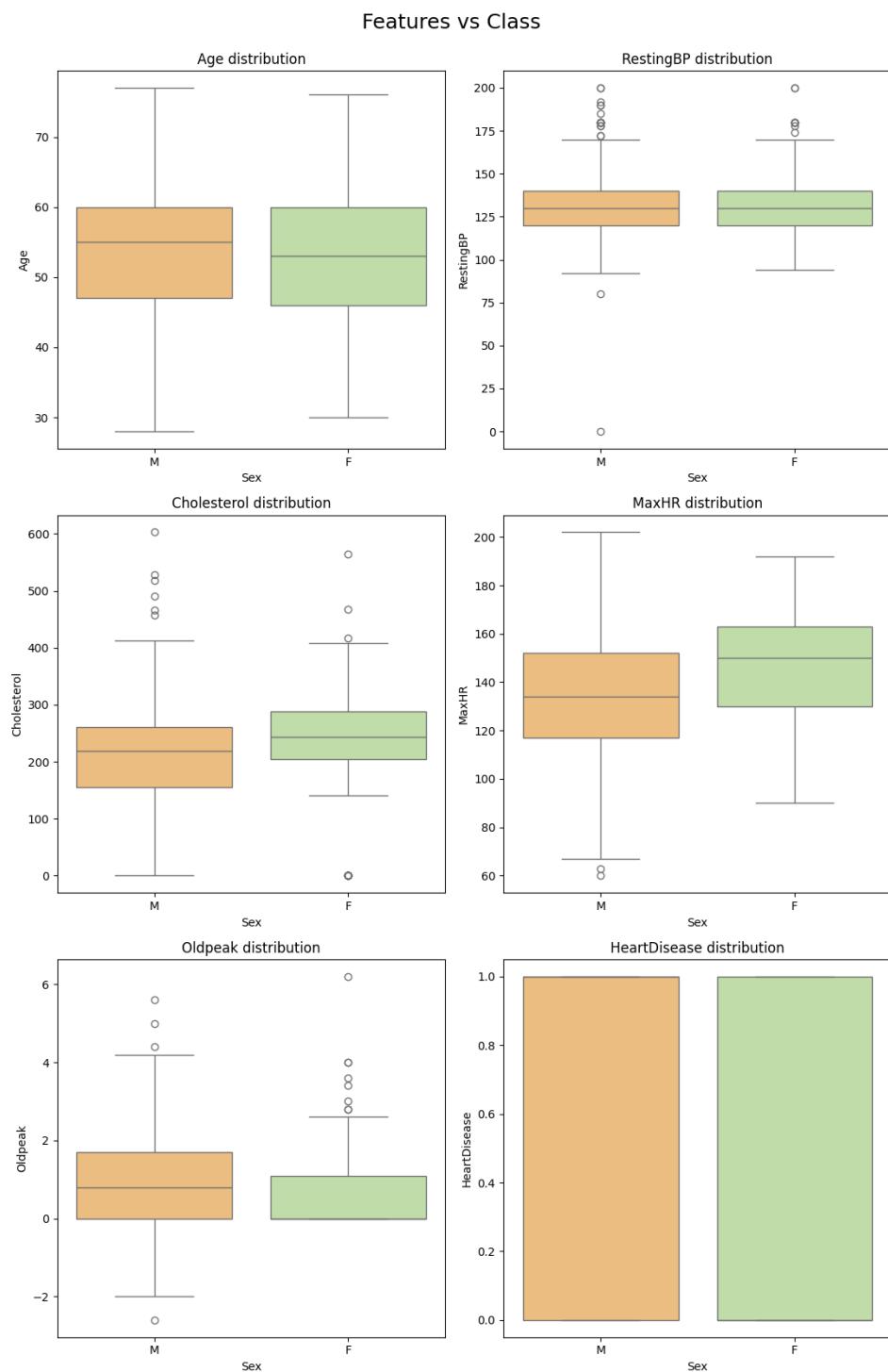


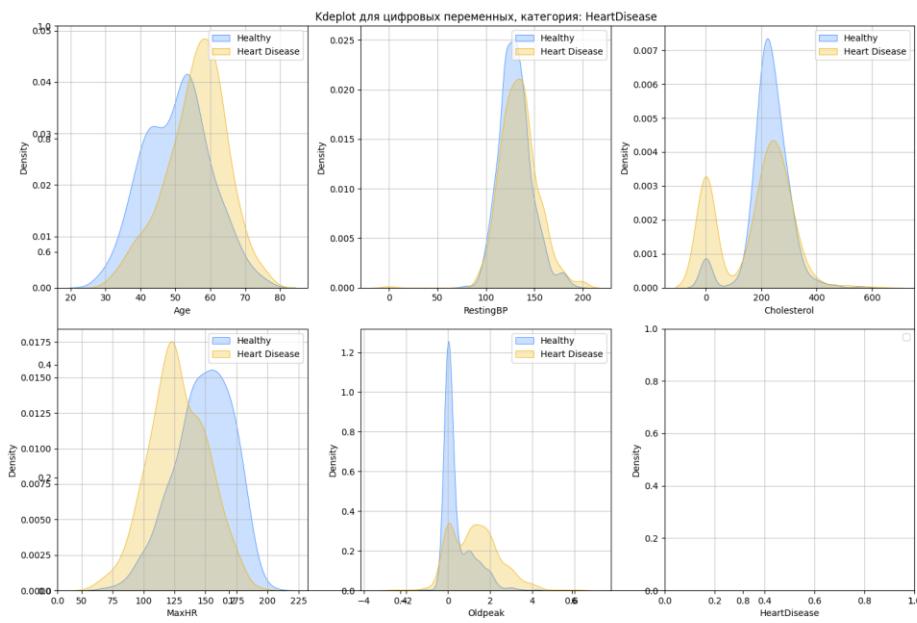


Таблица №5

Analysis: MaxHR by Geschlecht, Alter								
		N	Min	Max	Range	Mean	Median	Std
Female	0 - 29 years	4	160	202	42	179	178	18
	30 - <39 years	57	98	187	89	153	154	23
	40 - <65 years	580	60	195	135	134	134	25
	65 - <75 years	78	67	174	107	123	124	21
	75+ years	6	108	162	54	121	112	21
Male	30 - <39 years	19	129	192	63	164	166	19
	40 - <65 years	155	90	180	90	145	148	21
	65 - <75 years	18	90	172	82	138	145	23
	75+ years	1	116	116	0	116	116	
Total	0 - 29 years	4	160	202	42	179	178	18
	30 - <39 years	76	98	192	94	156	156	22
	40 - <65 years	735	60	195	135	136	138	25
	65 - <75 years	96	67	174	107	126	125	22
	75+ years	7	108	162	54	120	112	19

© Dr. Alexander Wagner. Все права охраняются законом

График №4. Гистограммы распределения для всех числовых переменных по классу Заболевание (HeartDisease)





## Таблица №6

<i>Analysis: Oldpeak by Geschlecht, Alter</i>		N	Min	Max	Range	Mean	Median	Std
Female	<b>30 - &lt;39 years</b>	19	0	4	4	2	1	1
	<b>40 - &lt;65 years</b>	356	0	6	6	2	2	1
	<b>65 - &lt;75 years</b>	63	0	4	4	2	2	1
	<b>75+ years</b>	4	2	4	2	3	3	1
Male	<b>30 - &lt;39 years</b>	4	1	3	2	1	1	1
	<b>40 - &lt;65 years</b>	77	0	6	6	1	1	1
	<b>65 - &lt;75 years</b>	13	0	3	2	1	1	1
	<b>75+ years</b>	1	1	1	0	1	1	
Total	<b>30 - &lt;39 years</b>	23	0	4	4	2	1	1
	<b>40 - &lt;65 years</b>	433	0	6	6	2	2	1
	<b>65 - &lt;75 years</b>	76	0	4	4	1	2	1
	<b>75+ years</b>	5	1	4	2	2	2	1

© Dr. Alexander Wagner. Все права охраняются законом





## График №5. Гистограммы распределения для всех числовых переменных по классу пол (Sex)

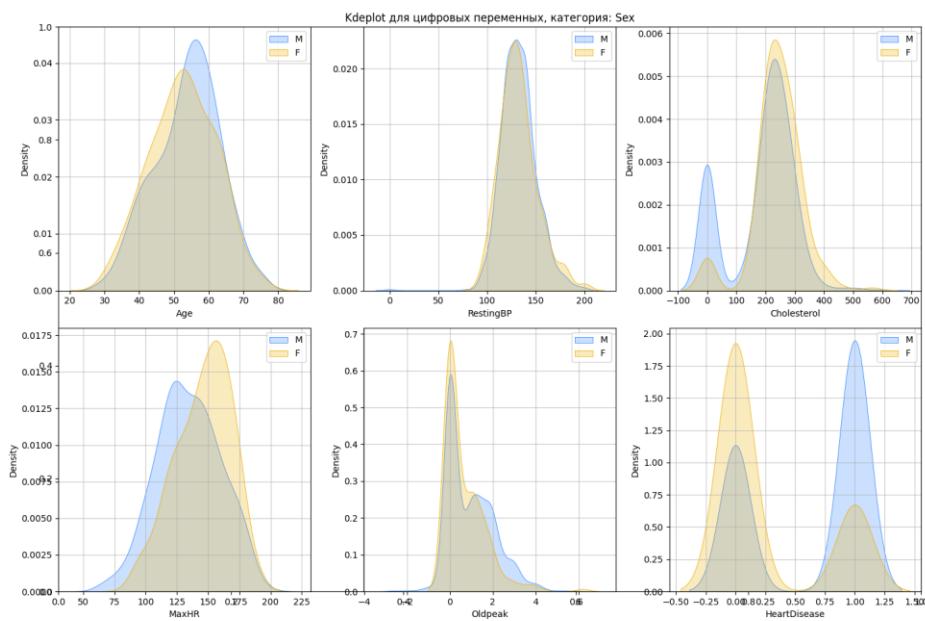


Таблица №7

Age	Healthy			Heart Disease			Всего			
	Mean	N	%	Mean	N	%	Mean	N	%	
Female	0 - 29 years	28.75	4	100.00			0.00	28.75	4	100.00
	30 - <39 years	36.68	34	59.65	36.17	23	40.35	36.47	57	100.00
	40 - <65 years	51.41	213	36.72	54.73	367	63.28	53.51	580	100.00
	65 - <75 years	67.80	15	19.23	68.10	63	80.77	68.04	78	100.00
	75+ years	75.00	1	16.67	76.00	5	83.33	75.83	6	100.00
	Всего	50.20	267	36.83	55.87	458	63.17	53.78	725	100.00
Male	0 - 29 years									
	30 - <39 years	35.56	16	84.21	36.33	3	15.79	35.68	19	100.00
	40 - <65 years	51.12	112	72.26	56.51	43	27.74	52.61	155	100.00
	65 - <75 years	68.00	14	77.78	67.50	4	22.22	67.89	18	100.00
	75+ years	76.00	1	100.00			0.00	76.00	1	100.00
	Всего	51.20	143	74.09	56.18	50	25.91	52.49	193	100.00
Total	0 - 29 years	28.75	4	100.00			0.00	28.75	4	100.00
	30 - <39 years	36.32	50	65.79	36.19	26	34.21	36.28	76	100.00
	40 - <65 years	51.31	325	44.22	54.92	410	55.78	53.32	735	100.00
	65 - <75 years	67.90	29	30.21	68.06	67	69.79	68.01	96	100.00
	75+ years	75.50	2	28.57	76.00	5	71.43	75.86	7	100.00
	Всего	50.55	410	44.66	55.90	508	55.34	53.51	918	100.00



Age		Healthy			Heart Disease			Всего		
		Mean	N	%	Mean	N	%	Mean	N	%
Всего	0 - 29 years	28.75	8	100.00			0.00	28.75	8	100.00
	30 - <39 years	36.32	100	65.79	36.19	52	34.21	36.28	152	100.00
	40 - <65 years	51.31	650	44.22	54.92	820	55.78	53.32	1470	100.00
	65 - <75 years	67.90	58	30.21	68.06	134	69.79	68.01	192	100.00
	75+ years	75.50	4	28.57	76.00	10	71.43	75.86	14	100.00
	Всего	50.55	820	44.66	55.90	1016	55.34	53.51	1836	100.00

© Dr. Alexander Wagner. Все права охраняются законом



## График №6. Двумерное распределение переменной 'HeartDisease' по классам Sex и RestingBP

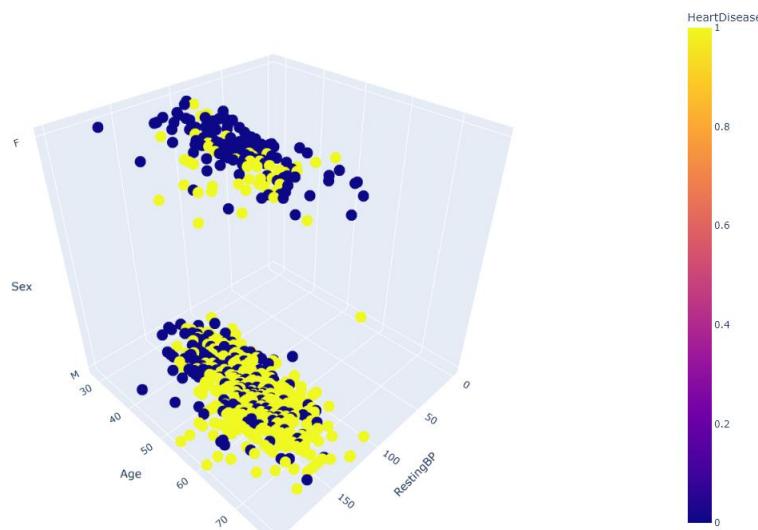


Таблица №8

Cholesterol		Healthy			Heart Disease			Всего			
		Mean	N	%	Mean	N	%	Mean	N	%	
Female	0 - 29 years	210.50	4	100.00				0.00	210.50	4	100.00
	30 - <39 years	240.91	33	67.35	248.13	16	32.65	243.27	49	100.00	
	40 - <65 years	231.16	199	44.42	249.12	249	55.58	241.14	448	100.00	
	65 - <75 years	239.45	11	19.30	243.33	46	80.70	242.58	57	100.00	
	75+ years	310.00	1	16.67	203.20	5	83.33	221.00	6	100.00	
	Всего	232.81	248	43.97	247.50	316	56.03	241.04	564	100.00	
Male	0 - 29 years										
	30 - <39 years	204.38	16	94.12	246.00	1	5.88	206.82	17	100.00	
	40 - <65 years	251.06	111	75.00	282.92	37	25.00	259.03	148	100.00	
	65 - <75 years	289.21	14	87.50	226.50	2	12.50	281.38	16	100.00	
	75+ years	197.00	1	100.00			0.00	197.00	1	100.00	
	Всего	249.18	142	78.02	279.18	40	21.98	255.77	182	100.00	
Total	0 - 29 years	210.50	4	100.00				0.00	210.50	4	100.00
	30 - <39 years	228.98	49	74.24	248.00	17	25.76	233.88	66	100.00	
	40 - <65 years	238.28	310	52.01	253.50	286	47.99	245.58	596	100.00	
	65 - <75 years	267.32	25	34.25	242.63	48	65.75	251.08	73	100.00	
	75+ years	253.50	2	28.57	203.20	5	71.43	217.57	7	100.00	
	Всего	238.77	390	52.28	251.06	356	47.72	244.64	746	100.00	



Cholesterol		Healthy			Heart Disease			Всего		
		Mean	N	%	Mean	N	%	Mean	N	%
<b>Всего</b>	<b>0 - 29 years</b>	210.50	8	100.00			0.00	210.50	8	100.00
	<b>30 - &lt;39 years</b>	228.98	98	74.24	248.00	34	25.76	233.88	132	100.00
	<b>40 - &lt;65 years</b>	238.28	620	52.01	253.50	572	47.99	245.58	1192	100.00
	<b>65 - &lt;75 years</b>	267.32	50	34.25	242.63	96	65.75	251.08	146	100.00
	<b>75+ years</b>	253.50	4	28.57	203.20	10	71.43	217.57	14	100.00
	<b>Всего</b>	238.77	780	52.28	251.06	712	47.72	244.64	1492	100.00

© Dr. Alexander Wagner. Все права охраняются законом



## График №7. График распределение переменной Cholesterol по Возрасту (Age)

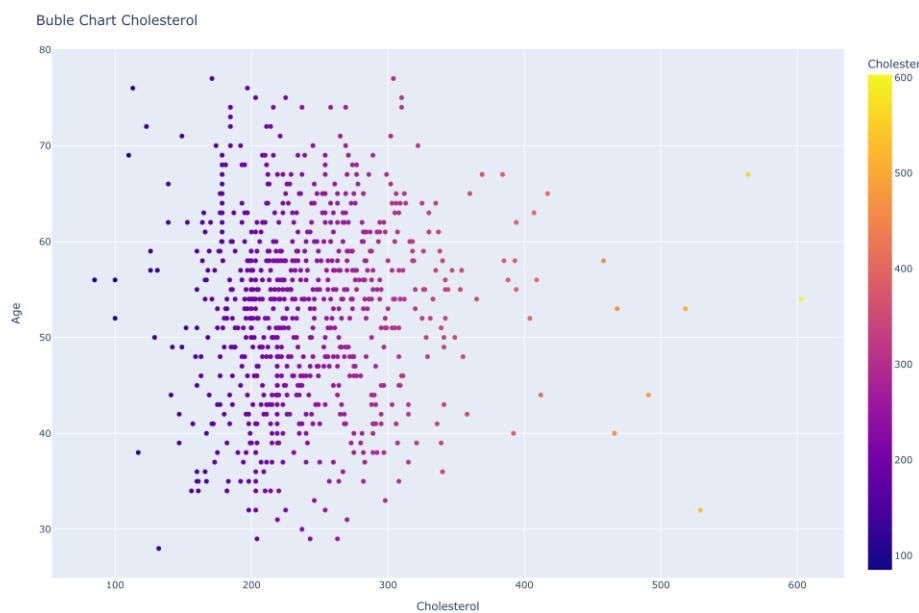


Таблица №9

RestingBP		Healthy			Heart Disease			Всего			
		Mean	N	%	Mean	N	%	Mean	N	%	
Female	0 - 29 years	130.00	4	100.00				0.00	130.00	4	100.00
	30 - <39 years	128.50	34	59.65	118.43	23	40.35	124.44	57	100.00	
	40 - <65 years	130.69	213	36.79	133.52	366	63.21	132.47	579	100.00	
	65 - <75 years	138.13	15	19.23	139.94	63	80.77	139.59	78	100.00	
	75+ years	160.00	1	16.67	131.80	5	83.33	136.50	6	100.00	
	Всего	130.93	267	36.88	133.62	457	63.12	132.63	724	100.00	
Male	0 - 29 years										
	30 - <39 years	123.94	16	84.21	105.00	3	15.79	120.95	19	100.00	
	40 - <65 years	128.64	112	72.26	142.84	43	27.74	132.58	155	100.00	
	65 - <75 years	134.71	14	77.78	160.75	4	22.22	140.50	18	100.00	
	75+ years	140.00	1	100.00			0.00	140.00	1	100.00	
	Всего	128.79	143	74.09	142.00	50	25.91	132.21	193	100.00	
Total	0 - 29 years	130.00	4	100.00				0.00	130.00	4	100.00
	30 - <39 years	127.04	50	65.79	116.88	26	34.21	123.57	76	100.00	
	40 - <65 years	129.98	325	44.28	134.50	409	55.72	132.50	734	100.00	
	65 - <75 years	136.48	29	30.21	141.18	67	69.79	139.76	96	100.00	
	75+ years	150.00	2	28.57	131.80	5	71.43	137.00	7	100.00	
	Всего	130.18	410	44.71	134.45	507	55.29	132.54	917	100.00	



RestingBP		Healthy			Heart Disease			Всего		
		Mean	N	%	Mean	N	%	Mean	N	%
<b>Всего</b>	<b>0 - 29 years</b>	130.00	8	100.00			0.00	130.00	8	100.00
	<b>30 - &lt;39 years</b>	127.04	100	65.79	116.88	52	34.21	123.57	152	100.00
	<b>40 - &lt;65 years</b>	129.98	650	44.28	134.50	818	55.72	132.50	1468	100.00
	<b>65 - &lt;75 years</b>	136.48	58	30.21	141.18	134	69.79	139.76	192	100.00
	<b>75+ years</b>	150.00	4	28.57	131.80	10	71.43	137.00	14	100.00
	<b>Всего</b>	130.18	820	44.71	134.45	1014	55.29	132.54	1834	100.00

© Dr. Alexander Wagner. Все права охраняются законом



## График №8. Гистограммы распределения для всех числовых переменных, представленные на одном графике

EDA на наборе данных показала нам, как каждая переменная связана с переменной отклика и как мы можем сделать нашу модель эффективной, используя различные методы EDA. Визуализации данных поднимают наше понимание набора данных на более высокий уровень, позволяя нам делать выводы.

Интеграция классификатора дерева принятия решений в наш анализ расширяет прогностические возможности нашей модели. В этом блоге мы не только изучили набор данных с помощью методов EDA, но и сделали еще один шаг вперед, внедрив модель машинного обучения. Такой целостный подход позволяет нам использовать сильные стороны как статистического анализа, так и прогнозного моделирования, способствуя более глубокому пониманию сложной динамики, связанной с экстремальными погодными явлениями.

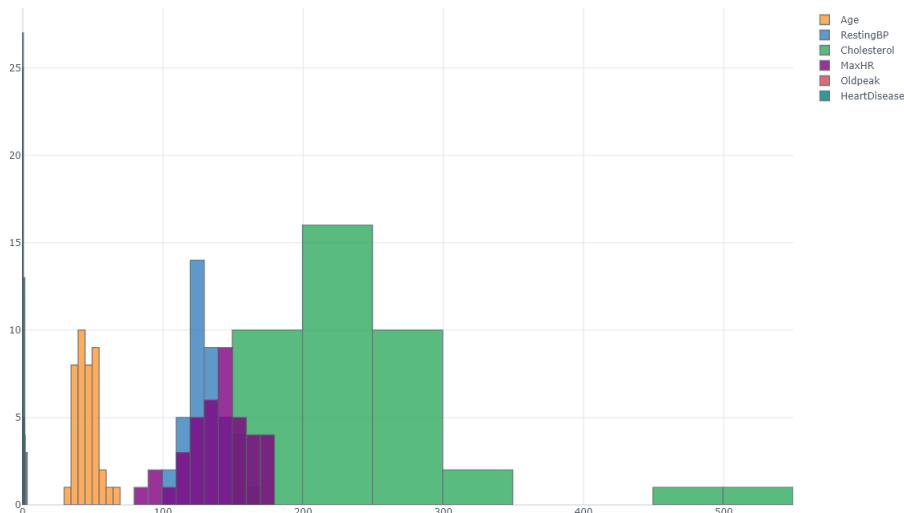


Таблица №10

MaxHR		Healthy			Heart Disease			Всего		
		Mean	N	%	Mean	N	%	Mean	N	%
Female	0 - 29 years	179.25	4	100.00			0.00	179.25	4	100.00
	30 - <39 years	159.53	34	59.65	143.91	23	40.35	153.23	57	100.00
	40 - <65 years	146.34	213	36.72	126.53	367	63.28	133.80	580	100.00
	65 - <75 years	133.60	15	19.23	120.65	63	80.77	123.14	78	100.00
	75+ years	112.00	1	16.67	122.40	5	83.33	120.67	6	100.00
	Всего	147.67	267	36.83	126.55	458	63.17	134.33	725	100.00



MaxHR		Healthy			Heart Disease			Всего		
		Mean	N	%	Mean	N	%	Mean	N	%
Male	<b>0 - 29 years</b>									
	<b>30 - &lt;39 years</b>	165.63	16	84.21	157.33	3	15.79	164.32	19	100.00
	<b>40 - &lt;65 years</b>	147.75	112	72.26	137.88	43	27.74	145.01	155	100.00
	<b>65 - &lt;75 years</b>	142.86	14	77.78	122.50	4	22.22	138.33	18	100.00
	<b>75+ years</b>	116.00	1	100.00			0.00	116.00	1	100.00
	<b>Всего</b>	149.05	143	74.09	137.82	50	25.91	146.14	193	100.00
Total	<b>0 - 29 years</b>	179.25	4	100.00			0.00	179.25	4	100.00
	<b>30 - &lt;39 years</b>	161.48	50	65.79	145.46	26	34.21	156.00	76	100.00
	<b>40 - &lt;65 years</b>	146.83	325	44.22	127.72	410	55.78	136.17	735	100.00
	<b>65 - &lt;75 years</b>	138.07	29	30.21	120.76	67	69.79	125.99	96	100.00
	<b>75+ years</b>	114.00	2	28.57	122.40	5	71.43	120.00	7	100.00
	<b>Всего</b>	148.15	410	44.66	127.66	508	55.34	136.81	918	100.00
Всего	<b>0 - 29 years</b>	179.25	8	100.00			0.00	179.25	8	100.00
	<b>30 - &lt;39 years</b>	161.48	100	65.79	145.46	52	34.21	156.00	152	100.00
	<b>40 - &lt;65 years</b>	146.83	650	44.22	127.72	820	55.78	136.17	1470	100.00
	<b>65 - &lt;75 years</b>	138.07	58	30.21	120.76	134	69.79	125.99	192	100.00
	<b>75+ years</b>	114.00	4	28.57	122.40	10	71.43	120.00	14	100.00
	<b>Всего</b>	148.15	820	44.66	127.66	1016	55.34	136.81	1836	100.00

© Dr. Alexander Wagner. Все права охраняются законом



## График №9. Матрица корреляции Пирсона для числовых переменных

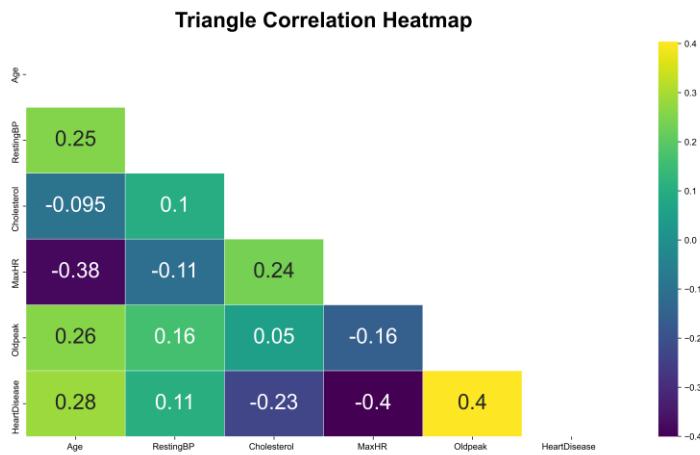


Таблица №11

Oldpeak		Healthy			Heart Disease			Всего		
		Mean	N	%	Mean	N	%	Mean	N	%
Female	30 - <39 years	1.70	5	26.32	1.53	14	73.68	1.57	19	100.00
	40 - <65 years	1.04	84	23.60	1.73	272	76.40	1.57	356	100.00
	65 - <75 years	0.76	12	19.05	1.76	51	80.95	1.57	63	100.00
	75+ years	2.00	1	25.00	2.83	3	75.00	2.63	4	100.00
	Всего	1.05	102	23.08	1.74	340	76.92	1.58	442	100.00
Male	30 - <39 years	1.05	2	50.00	1.90	2	50.00	1.47	4	100.00
	40 - <65 years	1.02	46	59.74	1.97	31	40.26	1.40	77	100.00
	65 - <75 years	1.13	11	84.62	1.00	2	15.38	1.11	13	100.00
	75+ years	1.10	1	100.00			0.00	1.10	1	100.00
	Всего	1.04	60	63.16	1.91	35	36.84	1.36	95	100.00
Total	30 - <39 years	1.51	7	30.43	1.58	16	69.57	1.56	23	100.00
	40 - <65 years	1.03	130	30.02	1.76	303	69.98	1.54	433	100.00
	65 - <75 years	0.93	23	30.26	1.73	53	69.74	1.49	76	100.00
	75+ years	1.55	2	40.00	2.83	3	60.00	2.32	5	100.00
	Всего	1.04	162	30.17	1.76	375	69.83	1.54	537	100.00
Всего	30 - <39 years	1.51	14	30.43	1.58	32	69.57	1.56	46	100.00
	40 - <65 years	1.03	260	30.02	1.76	606	69.98	1.54	866	100.00
	65 - <75 years	0.93	46	30.26	1.73	106	69.74	1.49	152	100.00
	75+ years	1.55	4	40.00	2.83	6	60.00	2.32	10	100.00
	Всего	1.04	324	30.17	1.76	750	69.83	1.54	1074	100.00

© Dr. Alexander Wagner. Все права охраняются законом





## График №10. Гистограммы распределения для всех числовых переменных, в виде субграфиков на одной панели

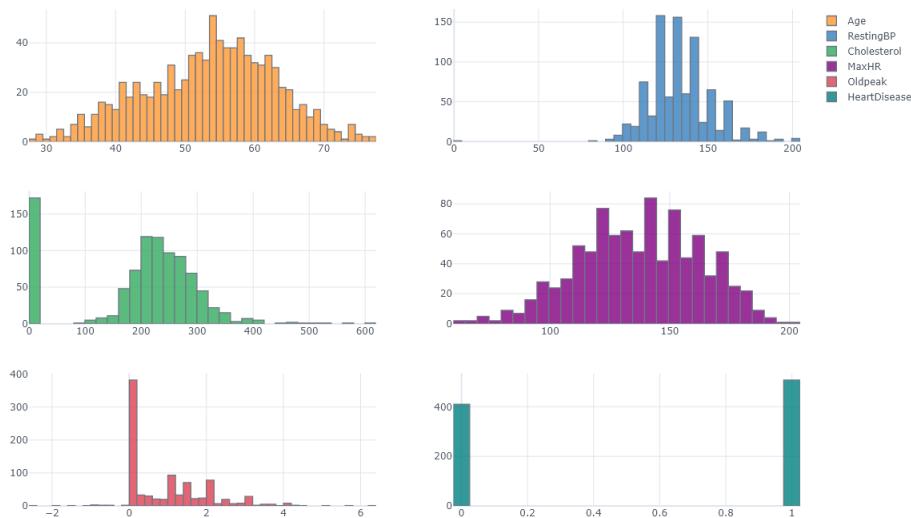




Таблица №12

ChestPainType			Healthy		Heart		Всего	
			N	%	N	%	N	%
ASY	Female	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	4	19.05	17	80.95	21	100.00
		<b>40 - &lt;65 years</b>	64	18.39	284	81.61	348	100.00
		<b>65 - &lt;75 years</b>	4	7.69	48	92.31	52	100.00
		<b>75+ years</b>	1	20.00	4	80.00	5	100.00
		<b>Всего</b>	73	17.14	353	82.86	426	100.00
	Male	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	3	50.00	3	50.00	6	100.00
		<b>40 - &lt;65 years</b>	26	44.07	33	55.93	59	100.00
		<b>65 - &lt;75 years</b>	2	40.00	3	60.00	5	100.00
		<b>75+ years</b>						
		<b>Всего</b>	31	44.29	39	55.71	70	100.00
	Total	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	7	25.93	20	74.07	27	100.00
		<b>40 - &lt;65 years</b>	90	22.11	317	77.89	407	100.00
		<b>65 - &lt;75 years</b>	6	10.53	51	89.47	57	100.00
		<b>75+ years</b>	1	20.00	4	80.00	5	100.00
		<b>Всего</b>	104	20.97	392	79.03	496	100.00
	Всего	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	14	25.93	40	74.07	54	100.00
		<b>40 - &lt;65 years</b>	180	22.11	634	77.89	814	100.00
		<b>65 - &lt;75 years</b>	12	10.53	102	89.47	114	100.00
		<b>75+ years</b>	2	20.00	8	80.00	10	100.00
		<b>Всего</b>	208	20.97	784	79.03	992	100.00



ChestPainType			Healthy		Heart		Всего	
			N	%	N	%	N	%
ATA	Female	<b>0 - 29 years</b>	4	100.00		0.00	4	100.00
		<b>30 - &lt;39 years</b>	15	88.24	2	11.76	17	100.00
		<b>40 - &lt;65 years</b>	73	82.02	16	17.98	89	100.00
		<b>65 - &lt;75 years</b>	1	33.33	2	66.67	3	100.00
		<b>75+ years</b>						
		<b>Всего</b>	93	82.30	20	17.70	113	100.00
	Male	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	6	100.00		0.00	6	100.00
		<b>40 - &lt;65 years</b>	48	92.31	4	7.69	52	100.00
		<b>65 - &lt;75 years</b>	2	100.00		0.00	2	100.00
		<b>75+ years</b>						
		<b>Всего</b>	56	93.33	4	6.67	60	100.00
	Total	<b>0 - 29 years</b>	4	100.00		0.00	4	100.00
		<b>30 - &lt;39 years</b>	21	91.30	2	8.70	23	100.00
		<b>40 - &lt;65 years</b>	121	85.82	20	14.18	141	100.00
		<b>65 - &lt;75 years</b>	3	60.00	2	40.00	5	100.00
		<b>75+ years</b>						
		<b>Всего</b>	149	86.13	24	13.87	173	100.00
	Всего	<b>0 - 29 years</b>	8	100.00		0.00	8	100.00
		<b>30 - &lt;39 years</b>	42	91.30	4	8.70	46	100.00
		<b>40 - &lt;65 years</b>	242	85.82	40	14.18	282	100.00
		<b>65 - &lt;75 years</b>	6	60.00	4	40.00	10	100.00
		<b>75+ years</b>						
		<b>Всего</b>	298	86.13	48	13.87	346	100.00



ChestPainType			Healthy		Heart		Всего	
			N	%	N	%	N	%
NAP	Female	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	14	93.33	1	6.67	15	100.00
		<b>40 - &lt;65 years</b>	63	53.39	55	46.61	118	100.00
		<b>65 - &lt;75 years</b>	7	43.75	9	56.25	16	100.00
		<b>75+ years</b>		0.00	1	100.00	1	100.00
		<b>Всего</b>	84	56.00	66	44.00	150	100.00
	Male	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	5	100.00		0.00	5	100.00
		<b>40 - &lt;65 years</b>	33	86.84	5	13.16	38	100.00
		<b>65 - &lt;75 years</b>	8	88.89	1	11.11	9	100.00
		<b>75+ years</b>	1	100.00		0.00	1	100.00
		<b>Всего</b>	47	88.68	6	11.32	53	100.00
	Total	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	19	95.00	1	5.00	20	100.00
		<b>40 - &lt;65 years</b>	96	61.54	60	38.46	156	100.00
		<b>65 - &lt;75 years</b>	15	60.00	10	40.00	25	100.00
		<b>75+ years</b>	1	50.00	1	50.00	2	100.00
		<b>Всего</b>	131	64.53	72	35.47	203	100.00
	Всего	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	38	95.00	2	5.00	40	100.00
		<b>40 - &lt;65 years</b>	192	61.54	120	38.46	312	100.00
		<b>65 - &lt;75 years</b>	30	60.00	20	40.00	50	100.00
		<b>75+ years</b>	2	50.00	2	50.00	4	100.00
		<b>Всего</b>	262	64.53	144	35.47	406	100.00



ChestPainType			Healthy		Heart		Всего	
			N	%	N	%	N	%
TA	Female	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	1	25.00	3	75.00	4	100.00
		<b>40 - &lt;65 years</b>	13	52.00	12	48.00	25	100.00
		<b>65 - &lt;75 years</b>	3	42.86	4	57.14	7	100.00
		<b>75+ years</b>						
		<b>Всего</b>	17	47.22	19	52.78	36	100.00
	Male	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	2	100.00		0.00	2	100.00
		<b>40 - &lt;65 years</b>	5	83.33	1	16.67	6	100.00
		<b>65 - &lt;75 years</b>	2	100.00		0.00	2	100.00
		<b>75+ years</b>						
		<b>Всего</b>	9	90.00	1	10.00	10	100.00
	Total	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	3	50.00	3	50.00	6	100.00
		<b>40 - &lt;65 years</b>	18	58.06	13	41.94	31	100.00
		<b>65 - &lt;75 years</b>	5	55.56	4	44.44	9	100.00
		<b>75+ years</b>						
		<b>Всего</b>	26	56.52	20	43.48	46	100.00
	Всего	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	6	50.00	6	50.00	12	100.00
		<b>40 - &lt;65 years</b>	36	58.06	26	41.94	62	100.00
		<b>65 - &lt;75 years</b>	10	55.56	8	44.44	18	100.00
		<b>75+ years</b>						
		<b>Всего</b>	52	56.52	40	43.48	92	100.00



ChestPainType			Healthy		Heart		Всего	
			N	%	N	%	N	%
Всего	Female	0 - 29 years	4	100.00		0.00	4	100.00
		30 - <39 years	34	59.65	23	40.35	57	100.00
		40 - <65 years	213	36.72	367	63.28	580	100.00
		65 - <75 years	15	19.23	63	80.77	78	100.00
		75+ years	1	16.67	5	83.33	6	100.00
	Male	Всего	267	36.83	458	63.17	725	100.00
		0 - 29 years						
		30 - <39 years	16	84.21	3	15.79	19	100.00
		40 - <65 years	112	72.26	43	27.74	155	100.00
		65 - <75 years	14	77.78	4	22.22	18	100.00
	Total	75+ years	1	100.00		0.00	1	100.00
		Всего	143	74.09	50	25.91	193	100.00
		0 - 29 years	4	100.00		0.00	4	100.00
		30 - <39 years	50	65.79	26	34.21	76	100.00
		40 - <65 years	325	44.22	410	55.78	735	100.00
	Всего	65 - <75 years	29	30.21	67	69.79	96	100.00
		75+ years	2	28.57	5	71.43	7	100.00
		Всего	410	44.66	508	55.34	918	100.00
		0 - 29 years	8	100.00		0.00	8	100.00
		30 - <39 years	100	65.79	52	34.21	152	100.00
	Всего	40 - <65 years	650	44.22	820	55.78	1470	100.00
		65 - <75 years	58	30.21	134	69.79	192	100.00
		75+ years	4	28.57	10	71.43	14	100.00
		Всего	820	44.66	1016	55.34	1836	100.00

© Dr. Alexander Wagner. Все права охраняются законом



## График №11. Распределение переменной Cholesterol по классу (Age, Sex, FastingBS) в виде 4 субграфиков на одной панели

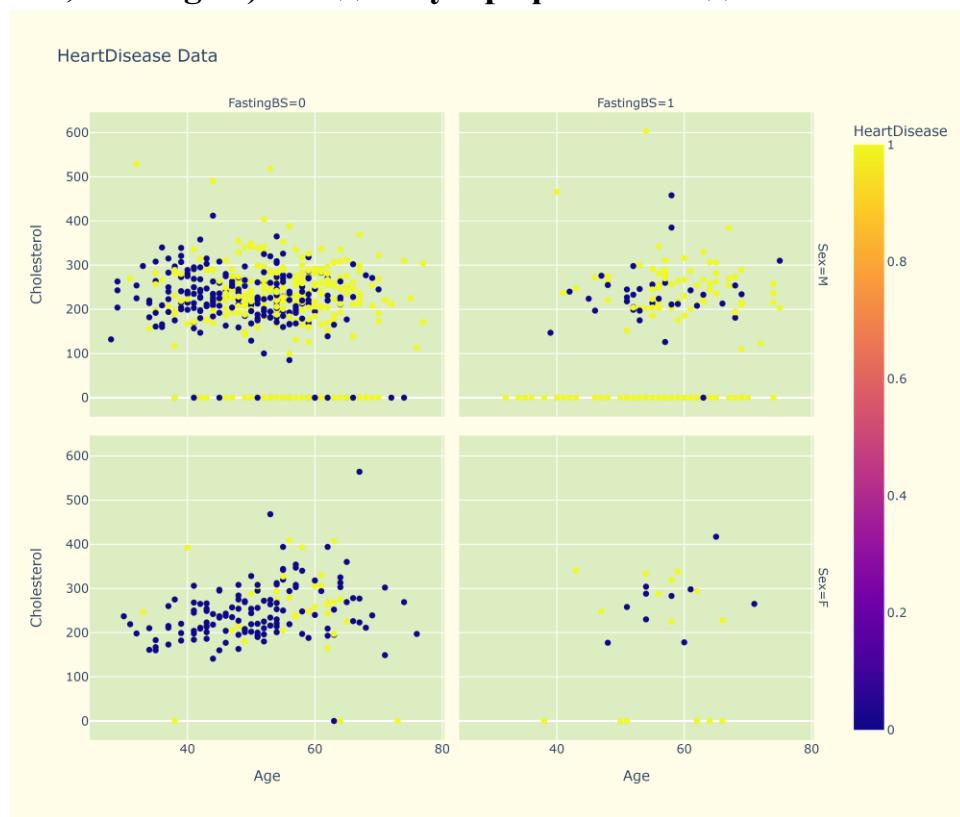




Таблица №13

ExerciseAngina			Healthy		Heart		Всего	
			N	%	N	%	N	%
N	Female	<b>0 - 29 years</b>	4	100.00		0.00	4	100.00
		<b>30 - &lt;39 years</b>	33	75.00	11	25.00	44	100.00
		<b>40 - &lt;65 years</b>	178	57.23	133	42.77	311	100.00
		<b>65 - &lt;75 years</b>	13	36.11	23	63.89	36	100.00
		<b>75+ years</b>		0.00	2	100.00	2	100.00
		<b>Всего</b>	228	57.43	169	42.57	397	100.00
	Male	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	16	88.89	2	11.11	18	100.00
		<b>40 - &lt;65 years</b>	97	84.35	18	15.65	115	100.00
		<b>65 - &lt;75 years</b>	13	81.25	3	18.75	16	100.00
		<b>75+ years</b>	1	100.00		0.00	1	100.00
		<b>Всего</b>	127	84.67	23	15.33	150	100.00
	Total	<b>0 - 29 years</b>	4	100.00		0.00	4	100.00
		<b>30 - &lt;39 years</b>	49	79.03	13	20.97	62	100.00
		<b>40 - &lt;65 years</b>	275	64.55	151	35.45	426	100.00
		<b>65 - &lt;75 years</b>	26	50.00	26	50.00	52	100.00
		<b>75+ years</b>	1	33.33	2	66.67	3	100.00
		<b>Всего</b>	355	64.90	192	35.10	547	100.00
	Всего	<b>0 - 29 years</b>	8	100.00		0.00	8	100.00
		<b>30 - &lt;39 years</b>	98	79.03	26	20.97	124	100.00
		<b>40 - &lt;65 years</b>	550	64.55	302	35.45	852	100.00
		<b>65 - &lt;75 years</b>	52	50.00	52	50.00	104	100.00
		<b>75+ years</b>	2	33.33	4	66.67	6	100.00
		<b>Всего</b>	710	64.90	384	35.10	1094	100.00



ExerciseAngina			Healthy		Heart		Всего	
			N	%	N	%	N	%
Y	Female	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	1	7.69	12	92.31	13	100.00
		<b>40 - &lt;65 years</b>	35	13.01	234	86.99	269	100.00
		<b>65 - &lt;75 years</b>	2	4.76	40	95.24	42	100.00
		<b>75+ years</b>	1	25.00	3	75.00	4	100.00
		<b>Всего</b>	39	11.89	289	88.11	328	100.00
	Male	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>		0.00	1	100.00	1	100.00
		<b>40 - &lt;65 years</b>	15	37.50	25	62.50	40	100.00
		<b>65 - &lt;75 years</b>	1	50.00	1	50.00	2	100.00
		<b>75+ years</b>						
		<b>Всего</b>	16	37.21	27	62.79	43	100.00
	Total	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	1	7.14	13	92.86	14	100.00
		<b>40 - &lt;65 years</b>	50	16.18	259	83.82	309	100.00
		<b>65 - &lt;75 years</b>	3	6.82	41	93.18	44	100.00
		<b>75+ years</b>	1	25.00	3	75.00	4	100.00
		<b>Всего</b>	55	14.82	316	85.18	371	100.00
	Всего	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	2	7.14	26	92.86	28	100.00
		<b>40 - &lt;65 years</b>	100	16.18	518	83.82	618	100.00
		<b>65 - &lt;75 years</b>	6	6.82	82	93.18	88	100.00
		<b>75+ years</b>	2	25.00	6	75.00	8	100.00
		<b>Всего</b>	110	14.82	632	85.18	742	100.00



ExerciseAngina			Healthy		Heart		Всего	
			N	%	N	%	N	%
Всего	Female	0 - 29 years	4	100.00		0.00	4	100.00
		30 - <39 years	34	59.65	23	40.35	57	100.00
		40 - <65 years	213	36.72	367	63.28	580	100.00
		65 - <75 years	15	19.23	63	80.77	78	100.00
		75+ years	1	16.67	5	83.33	6	100.00
	Male	Всего	267	36.83	458	63.17	725	100.00
		0 - 29 years						
		30 - <39 years	16	84.21	3	15.79	19	100.00
		40 - <65 years	112	72.26	43	27.74	155	100.00
		65 - <75 years	14	77.78	4	22.22	18	100.00
	Total	75+ years	1	100.00		0.00	1	100.00
		Всего	143	74.09	50	25.91	193	100.00
		0 - 29 years	4	100.00		0.00	4	100.00
		30 - <39 years	50	65.79	26	34.21	76	100.00
		40 - <65 years	325	44.22	410	55.78	735	100.00
	Всего	65 - <75 years	29	30.21	67	69.79	96	100.00
		75+ years	2	28.57	5	71.43	7	100.00
		Всего	410	44.66	508	55.34	918	100.00
		0 - 29 years	8	100.00		0.00	8	100.00
		30 - <39 years	100	65.79	52	34.21	152	100.00
	Всего	40 - <65 years	650	44.22	820	55.78	1470	100.00
		65 - <75 years	58	30.21	134	69.79	192	100.00
		75+ years	4	28.57	10	71.43	14	100.00
		Всего	820	44.66	1016	55.34	1836	100.00

© Dr. Alexander Wagner. Все права охраняются законом



**График №12. Распределение по возрасту (Age) для переменных: 'Sex','ChestPainType','FastingBS','RestingECG','ExerciseAngina','ST\_Slope','HeartDisease' в форме Виалин-графиков**

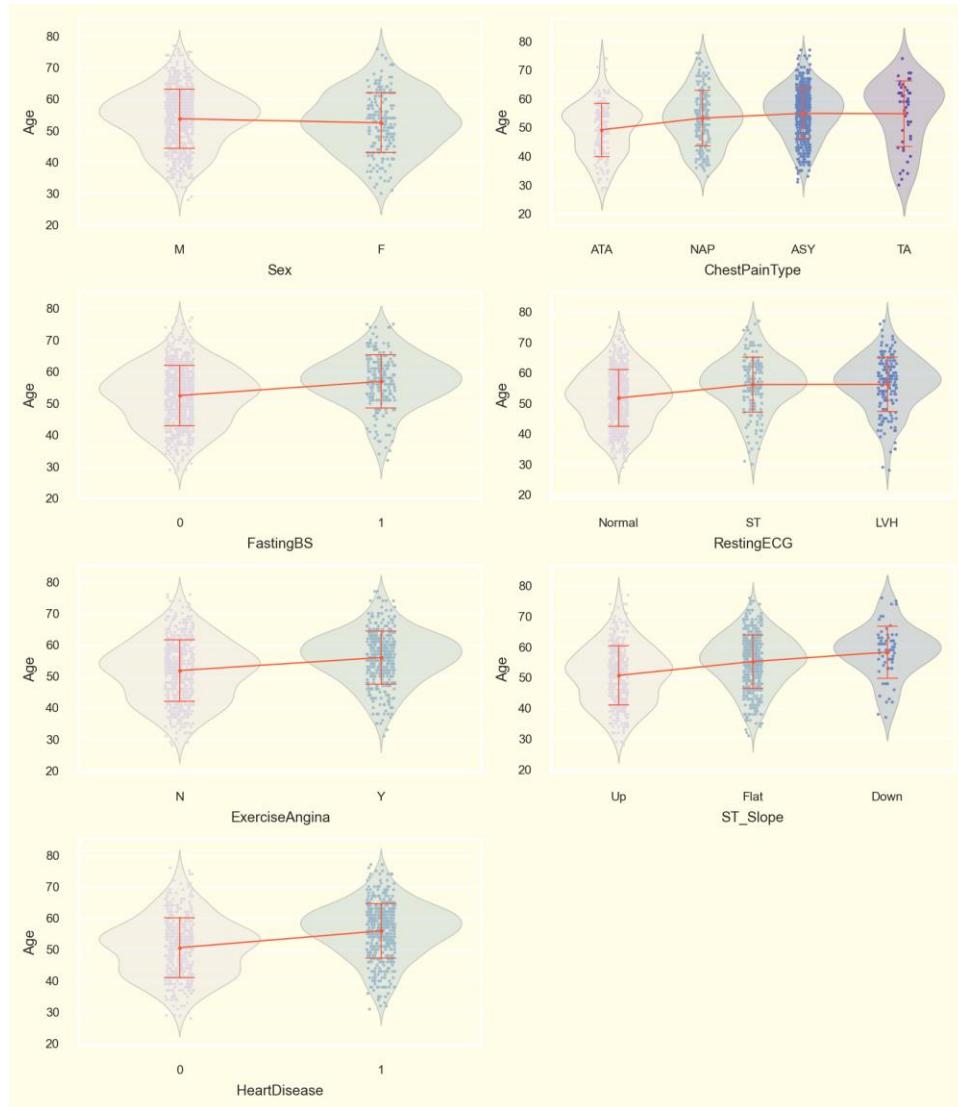




Таблица №14

RestingECG			Healthy		Heart		Всего	
			N	%	N	%	N	%
LVH	Female	<b>0 - 29 years</b>	2	100.00		0.00	2	100.00
		<b>30 - &lt;39 years</b>	4	80.00	1	20.00	5	100.00
		<b>40 - &lt;65 years</b>	36	32.73	74	67.27	110	100.00
		<b>65 - &lt;75 years</b>	7	31.82	15	68.18	22	100.00
		<b>75+ years</b>		0.00	2	100.00	2	100.00
		<b>Всего</b>	49	34.75	92	65.25	141	100.00
	Male	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>						
		<b>40 - &lt;65 years</b>	26	66.67	13	33.33	39	100.00
		<b>65 - &lt;75 years</b>	7	87.50	1	12.50	8	100.00
		<b>75+ years</b>						
		<b>Всего</b>	33	70.21	14	29.79	47	100.00
	Total	<b>0 - 29 years</b>	2	100.00		0.00	2	100.00
		<b>30 - &lt;39 years</b>	4	80.00	1	20.00	5	100.00
		<b>40 - &lt;65 years</b>	62	41.61	87	58.39	149	100.00
		<b>65 - &lt;75 years</b>	14	46.67	16	53.33	30	100.00
		<b>75+ years</b>		0.00	2	100.00	2	100.00
		<b>Всего</b>	82	43.62	106	56.38	188	100.00
	Всего	<b>0 - 29 years</b>	4	100.00		0.00	4	100.00
		<b>30 - &lt;39 years</b>	8	80.00	2	20.00	10	100.00
		<b>40 - &lt;65 years</b>	124	41.61	174	58.39	298	100.00
		<b>65 - &lt;75 years</b>	28	46.67	32	53.33	60	100.00
		<b>75+ years</b>		0.00	4	100.00	4	100.00
		<b>Всего</b>	164	43.62	212	56.38	376	100.00



RestingECG			Healthy		Heart		Всего	
			N	%	N	%	N	%
Normal	Female	<b>0 - 29 years</b>	2	100.00		0.00	2	100.00
		<b>30 - &lt;39 years</b>	26	54.17	22	45.83	48	100.00
		<b>40 - &lt;65 years</b>	143	41.33	203	58.67	346	100.00
		<b>65 - &lt;75 years</b>	6	16.67	30	83.33	36	100.00
		<b>75+ years</b>	1	50.00	1	50.00	2	100.00
		<b>Всего</b>	178	41.01	256	58.99	434	100.00
	Male	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	11	78.57	3	21.43	14	100.00
		<b>40 - &lt;65 years</b>	71	74.74	24	25.26	95	100.00
		<b>65 - &lt;75 years</b>	7	77.78	2	22.22	9	100.00
		<b>75+ years</b>						
		<b>Всего</b>	89	75.42	29	24.58	118	100.00
	Total	<b>0 - 29 years</b>	2	100.00		0.00	2	100.00
		<b>30 - &lt;39 years</b>	37	59.68	25	40.32	62	100.00
		<b>40 - &lt;65 years</b>	214	48.53	227	51.47	441	100.00
		<b>65 - &lt;75 years</b>	13	28.89	32	71.11	45	100.00
		<b>75+ years</b>	1	50.00	1	50.00	2	100.00
		<b>Всего</b>	267	48.37	285	51.63	552	100.00
	Всего	<b>0 - 29 years</b>	4	100.00		0.00	4	100.00
		<b>30 - &lt;39 years</b>	74	59.68	50	40.32	124	100.00
		<b>40 - &lt;65 years</b>	428	48.53	454	51.47	882	100.00
		<b>65 - &lt;75 years</b>	26	28.89	64	71.11	90	100.00
		<b>75+ years</b>	2	50.00	2	50.00	4	100.00
		<b>Всего</b>	534	48.37	570	51.63	1104	100.00



RestingECG			Healthy		Heart		Всего	
			N	%	N	%	N	%
ST	Female	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	4	100.00		0.00	4	100.00
		<b>40 - &lt;65 years</b>	34	27.42	90	72.58	124	100.00
		<b>65 - &lt;75 years</b>	2	10.00	18	90.00	20	100.00
		<b>75+ years</b>		0.00	2	100.00	2	100.00
		<b>Всего</b>	40	26.67	110	73.33	150	100.00
	Male	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	5	100.00		0.00	5	100.00
		<b>40 - &lt;65 years</b>	15	71.43	6	28.57	21	100.00
		<b>65 - &lt;75 years</b>		0.00	1	100.00	1	100.00
		<b>75+ years</b>	1	100.00		0.00	1	100.00
		<b>Всего</b>	21	75.00	7	25.00	28	100.00
	Total	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	9	100.00		0.00	9	100.00
		<b>40 - &lt;65 years</b>	49	33.79	96	66.21	145	100.00
		<b>65 - &lt;75 years</b>	2	9.52	19	90.48	21	100.00
		<b>75+ years</b>	1	33.33	2	66.67	3	100.00
		<b>Всего</b>	61	34.27	117	65.73	178	100.00
	Всего	<b>0 - 29 years</b>						
		<b>30 - &lt;39 years</b>	18	100.00		0.00	18	100.00
		<b>40 - &lt;65 years</b>	98	33.79	192	66.21	290	100.00
		<b>65 - &lt;75 years</b>	4	9.52	38	90.48	42	100.00
		<b>75+ years</b>	2	33.33	4	66.67	6	100.00
		<b>Всего</b>	122	34.27	234	65.73	356	100.00

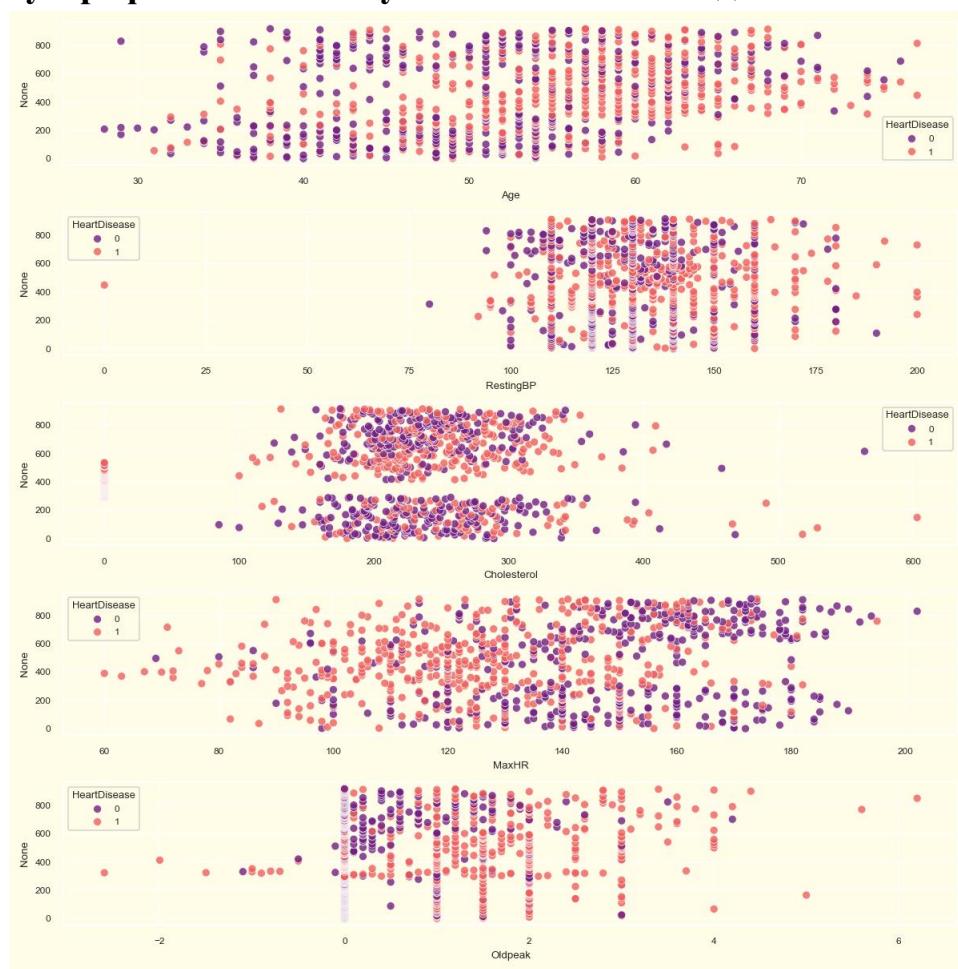


RestingECG			Healthy		Heart		Всего	
			N	%	N	%	N	%
Всего	Female	0 - 29 years	4	100.00		0.00	4	100.00
		30 - <39 years	34	59.65	23	40.35	57	100.00
		40 - <65 years	213	36.72	367	63.28	580	100.00
		65 - <75 years	15	19.23	63	80.77	78	100.00
		75+ years	1	16.67	5	83.33	6	100.00
		Всего	267	36.83	458	63.17	725	100.00
		0 - 29 years						
	Male	30 - <39 years	16	84.21	3	15.79	19	100.00
		40 - <65 years	112	72.26	43	27.74	155	100.00
		65 - <75 years	14	77.78	4	22.22	18	100.00
		75+ years	1	100.00		0.00	1	100.00
		Всего	143	74.09	50	25.91	193	100.00
		0 - 29 years						
		Total	0 - 29 years	4	100.00		0.00	4
	Всего	30 - <39 years	50	65.79	26	34.21	76	100.00
		40 - <65 years	325	44.22	410	55.78	735	100.00
		65 - <75 years	29	30.21	67	69.79	96	100.00
		75+ years	2	28.57	5	71.43	7	100.00
		Всего	410	44.66	508	55.34	918	100.00
		0 - 29 years						
		0 - 29 years	8	100.00		0.00	8	100.00
		30 - <39 years	100	65.79	52	34.21	152	100.00
		40 - <65 years	650	44.22	820	55.78	1470	100.00
		65 - <75 years	58	30.21	134	69.79	192	100.00
		75+ years	4	28.57	10	71.43	14	100.00
		Всего	820	44.66	1016	55.34	1836	100.00

© Dr. Alexander Wagner. Все права охраняются законом

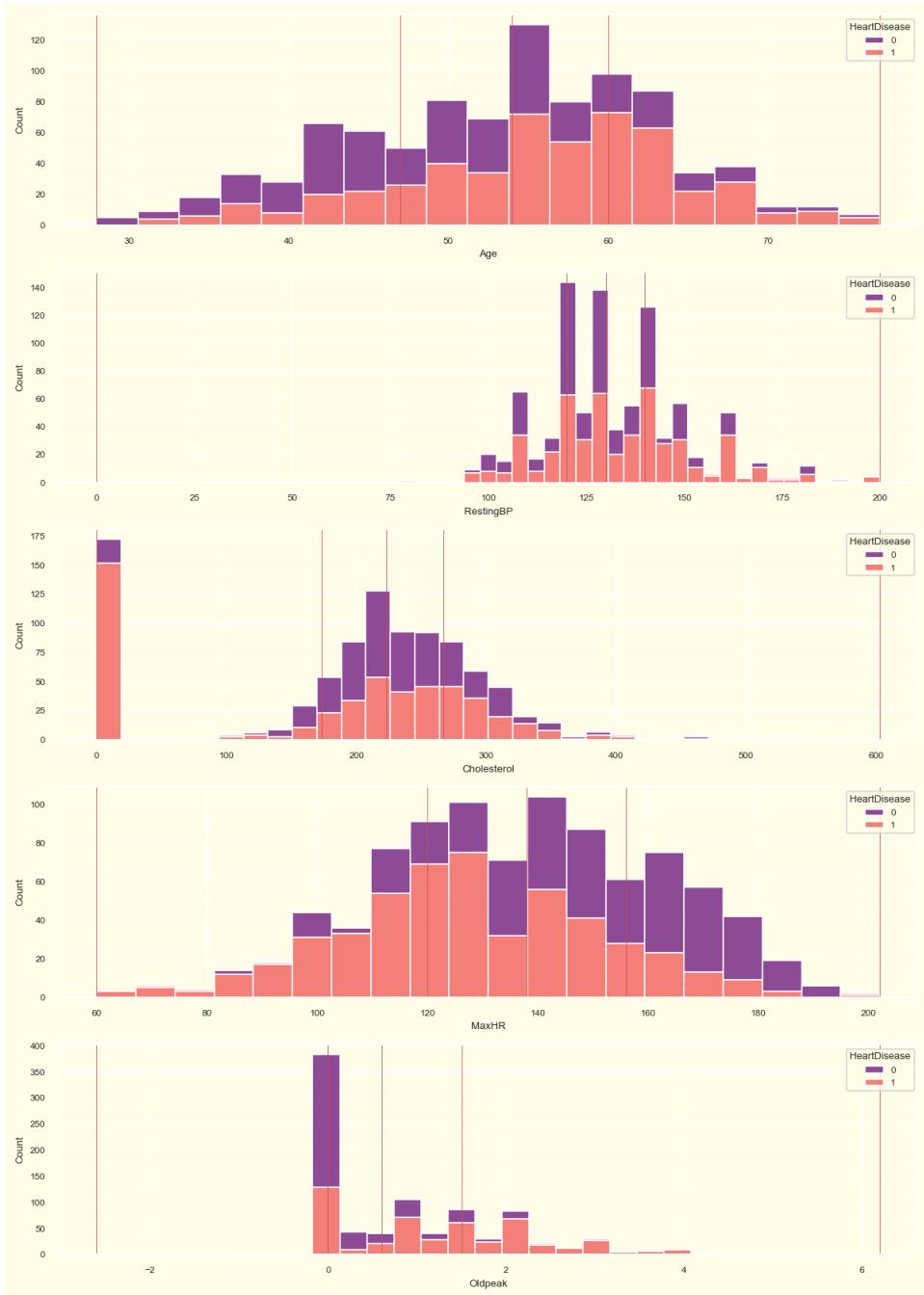


**График №13. Распределение всех числовых переменных в виде субграфиков по классу 'HeartDisease' на одной панели**



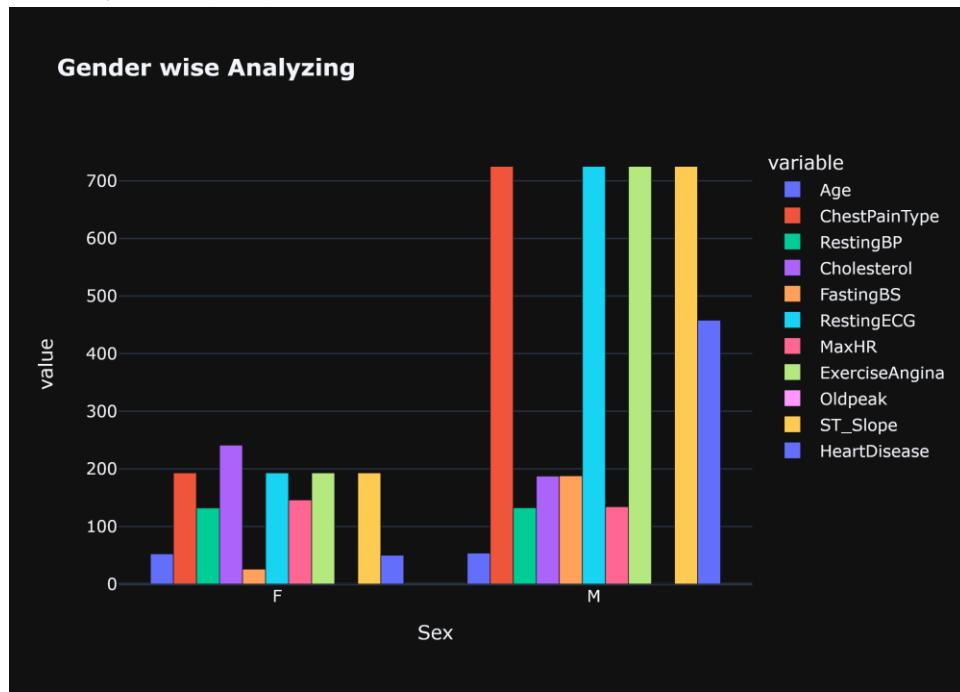


**График №14. Распределение всех числовых переменных в виде столбиковых диаграмм как субграфиков по классу 'HeartDisease' на одной панели**





**График №15. Распределение численности пациентов по всем переменным в виде столбиковых диаграмм как субграфиков по классу пол (Sex) на одной панели**





**График №16. Биполярное распределение переменной Cholesterol по возрасту 'Age'**

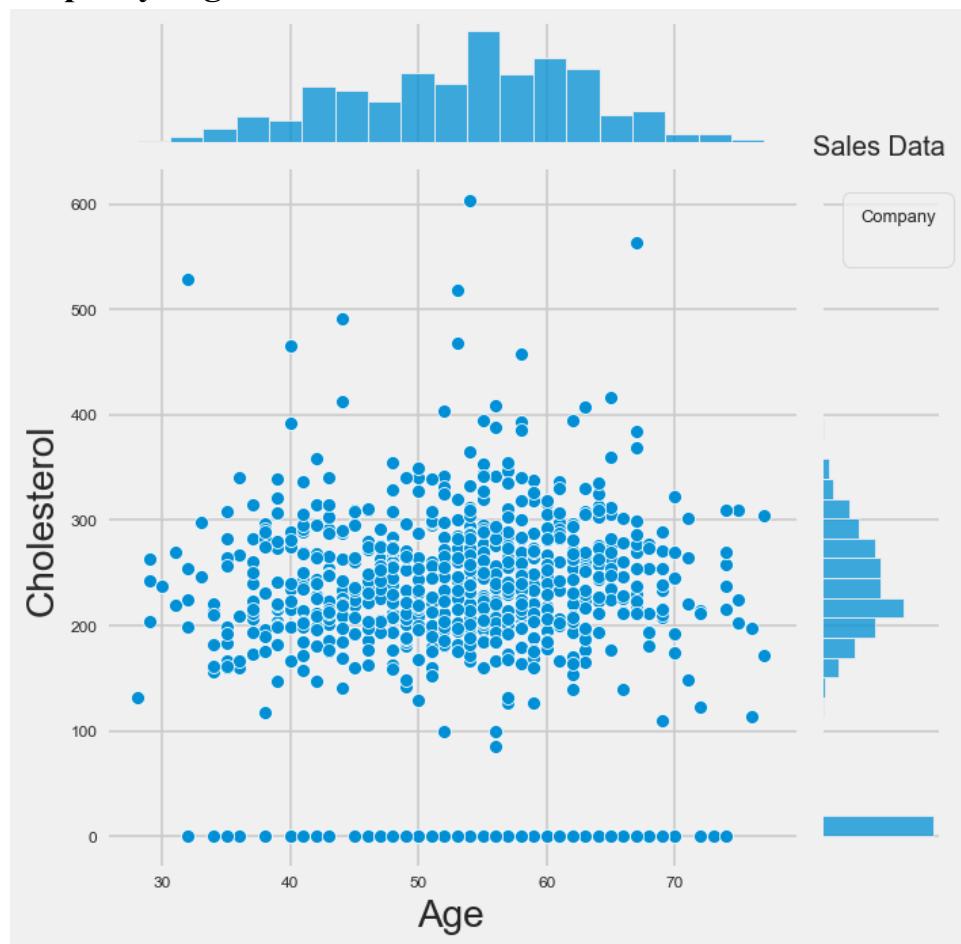
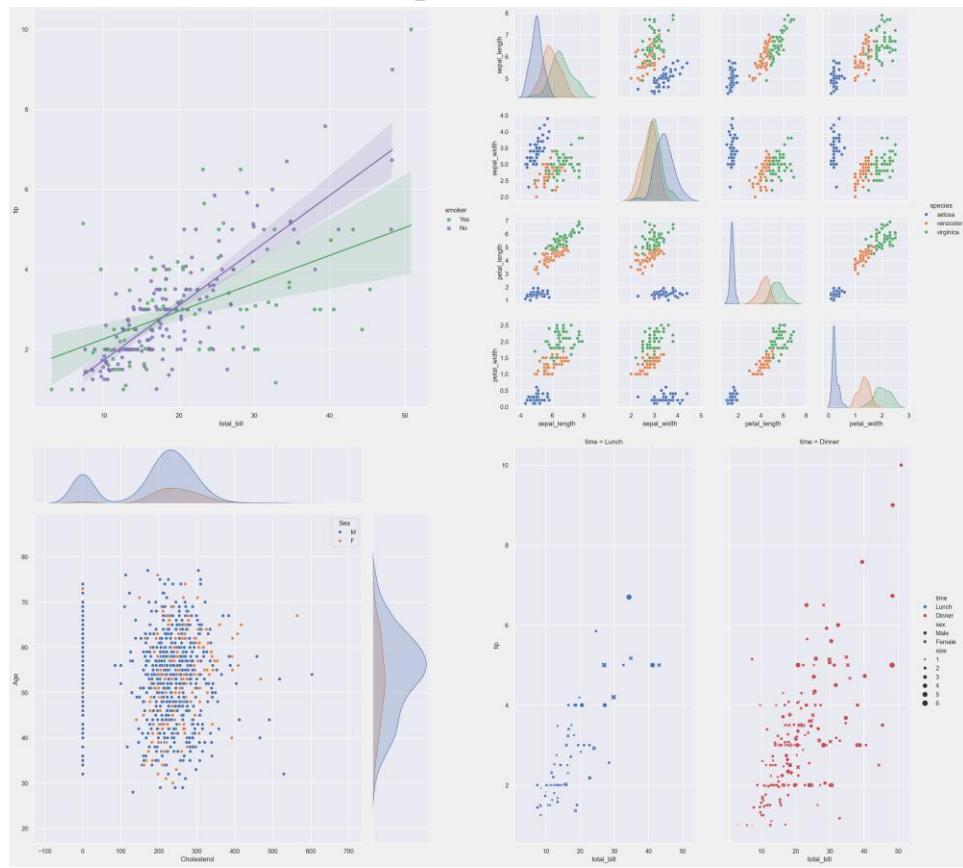






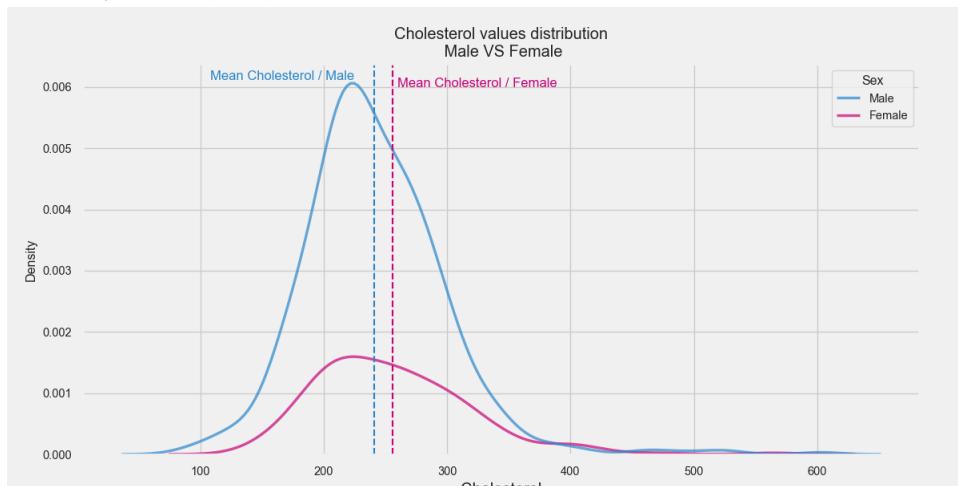
График №17. Биполярное распределение разного графического типа всех численных переменных на одной панели







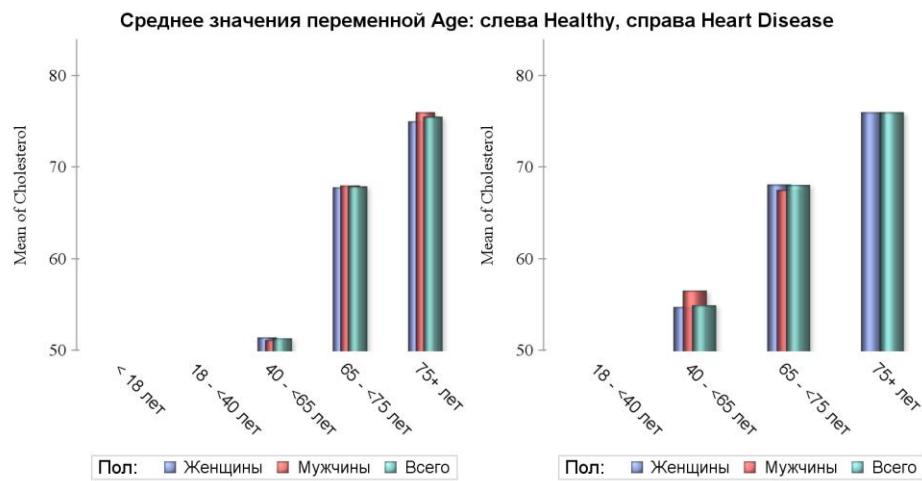
## График №18. Плотность распределения переменной Cholesterol по классу пол (Sex)







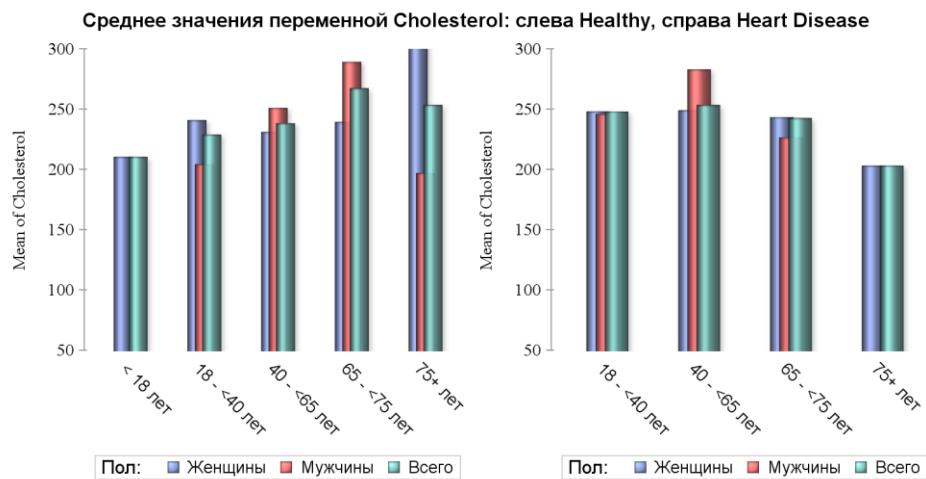
### График №19. Распределение переменной Age возрастным группам в виде столбиковых диаграмм по классу 'HeartDisease'







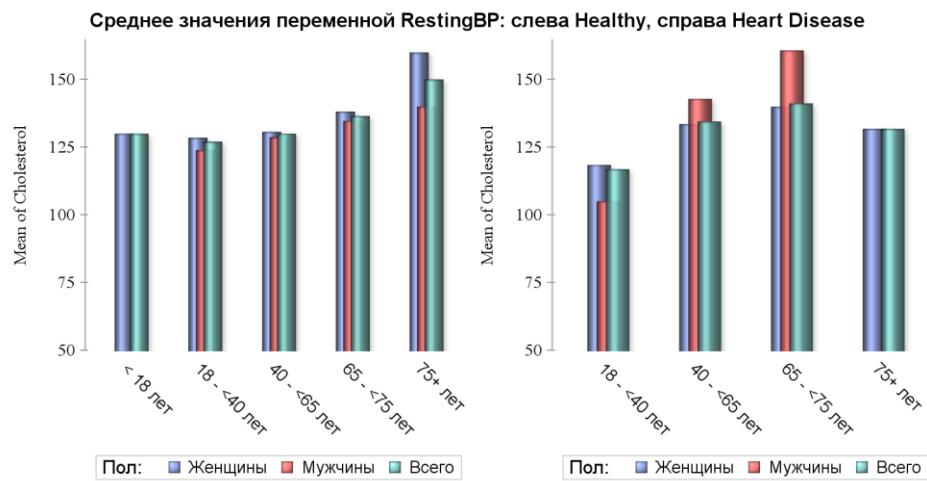
## График №20. Распределение переменной Cholesterol по возрастным группам в виде столбиковых диаграмм по классу 'HeartDisease'







## График №21. Распределение переменной RestingBP по возрастным группам в виде столбиковых диаграмм по классу 'HeartDisease'







## График №22. Распределение переменной MaxHR по возрастным группам в виде столбиковых диаграмм по классу 'HeartDisease'







### График №23. Распределение переменной Oldpeak по возрастным группам в виде столбиковых диаграмм по классу 'HeartDisease'







**График № 24. Торт-диаграмма распределения пациентов по классу переменной HeartDisease**

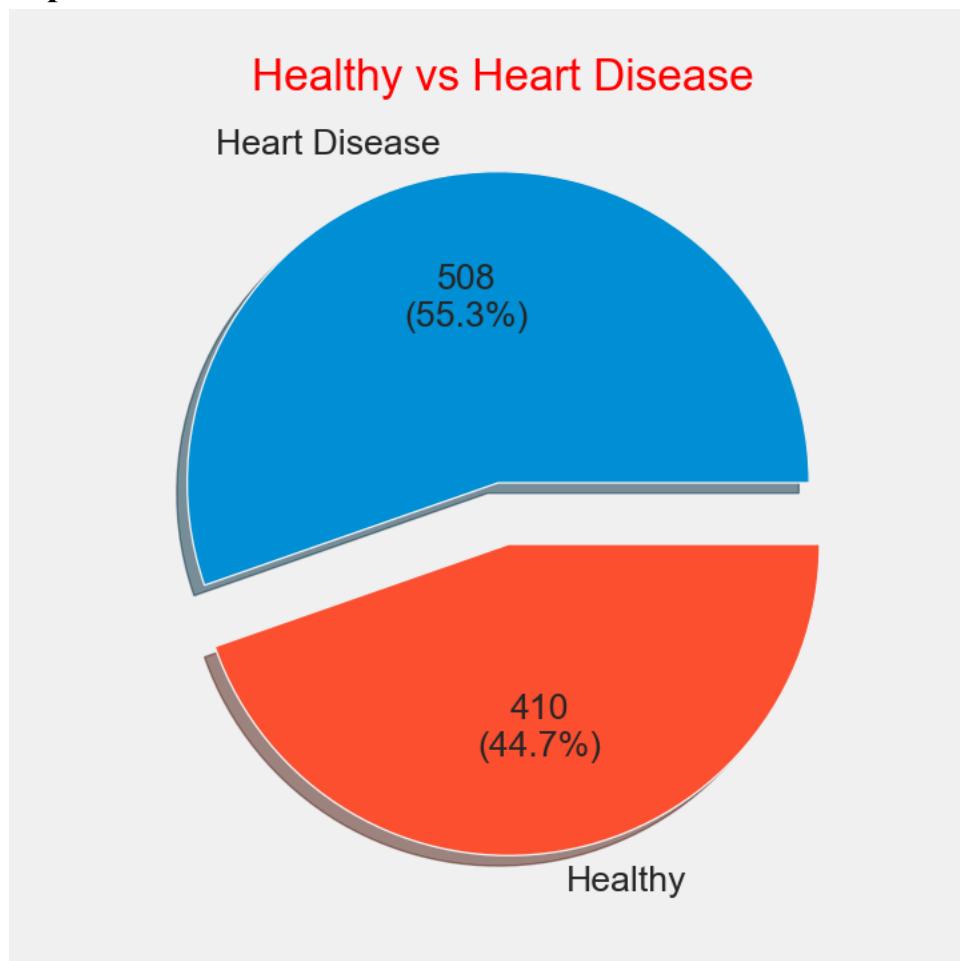
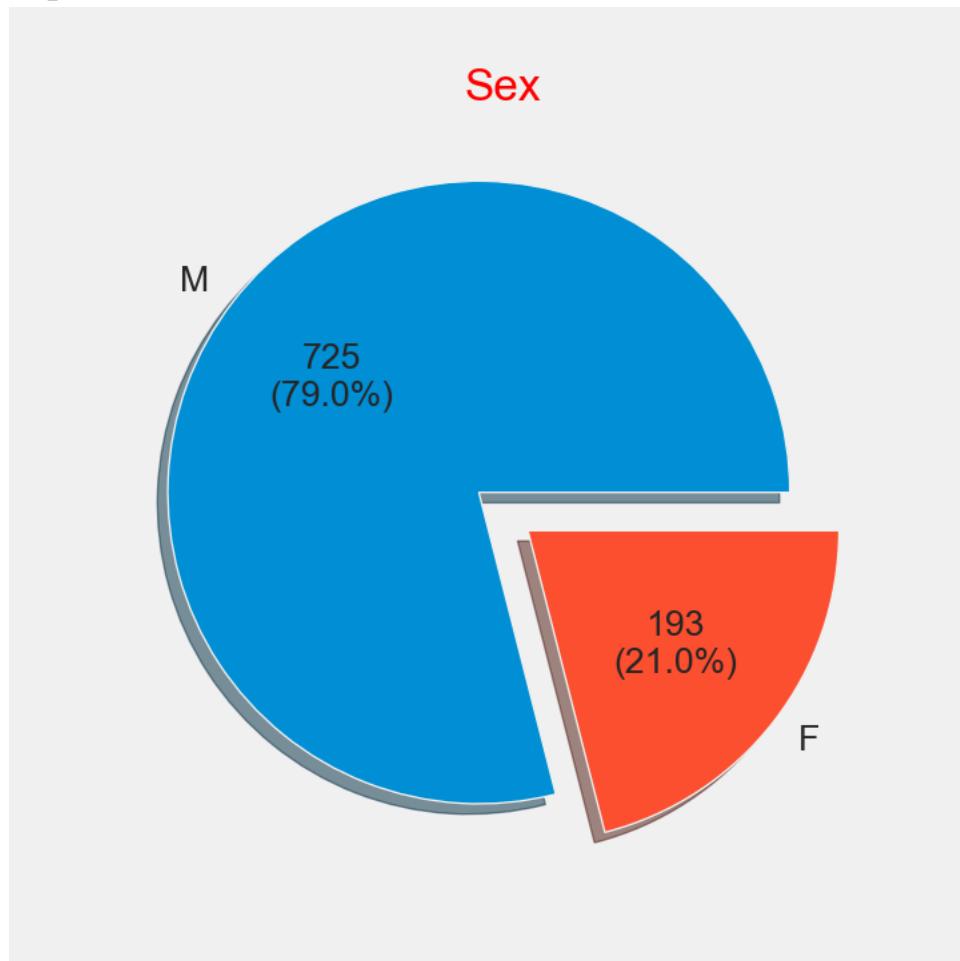






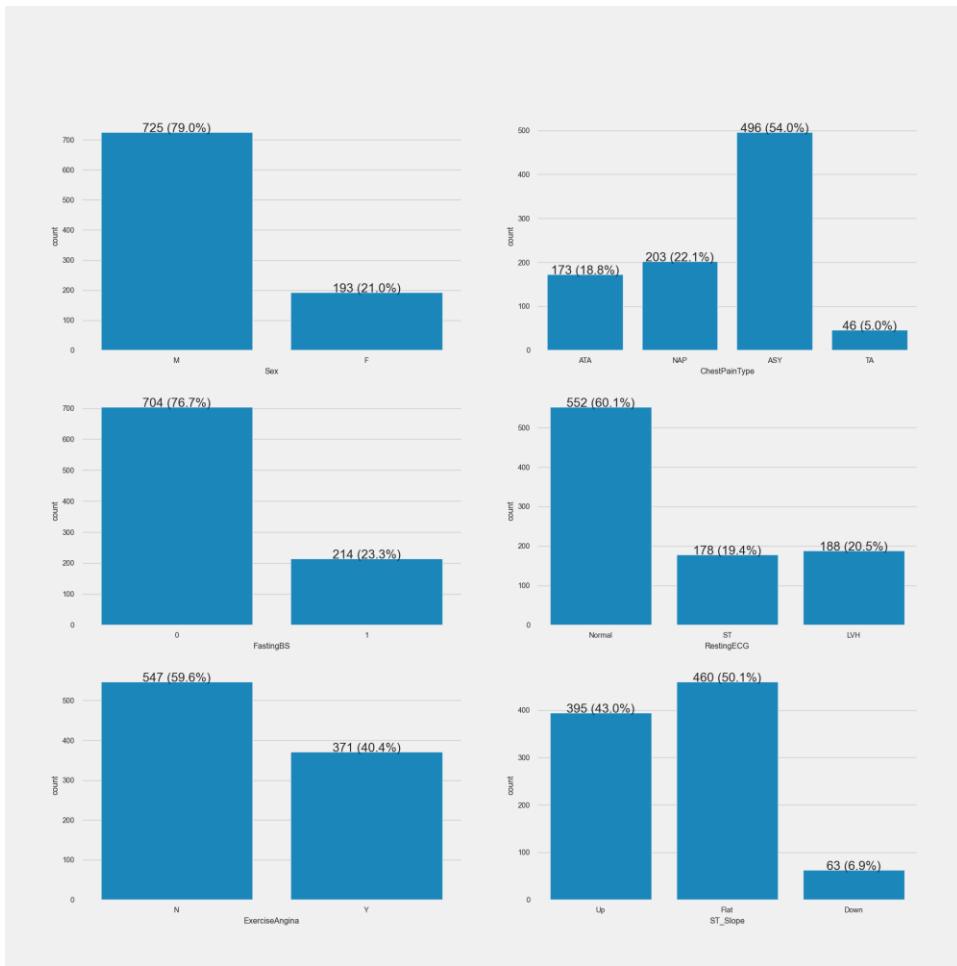
График №25. Торт-диаграмма распределения пациентов по классу-переменной Sex







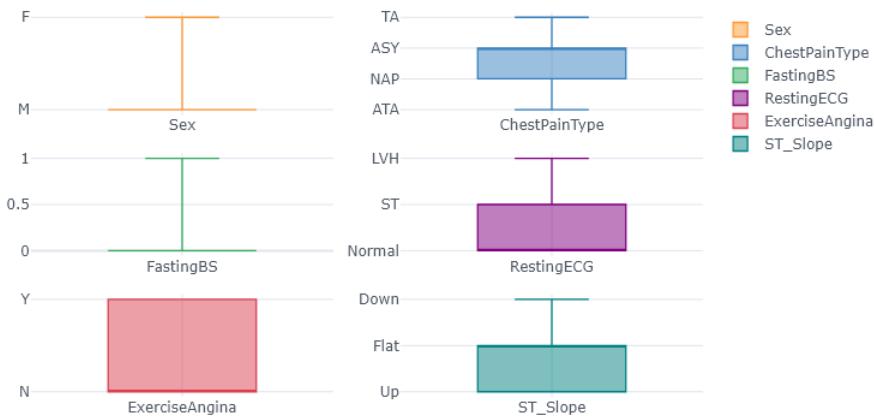
**График №26. Столбиковая диаграмма распределения переменных: 'Sex', 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST\_Slope', 'HeartDisease' по категориям**







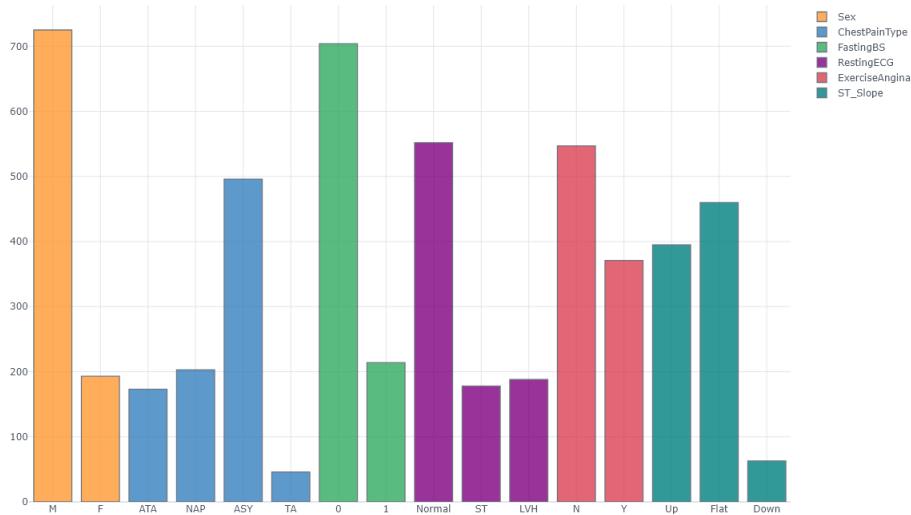
**График №27. Блок-бокс диаграмма распределения переменных: 'Sex', 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST\_Slope', 'HeartDisease' в виде субграфиков на одной панели по категориям**







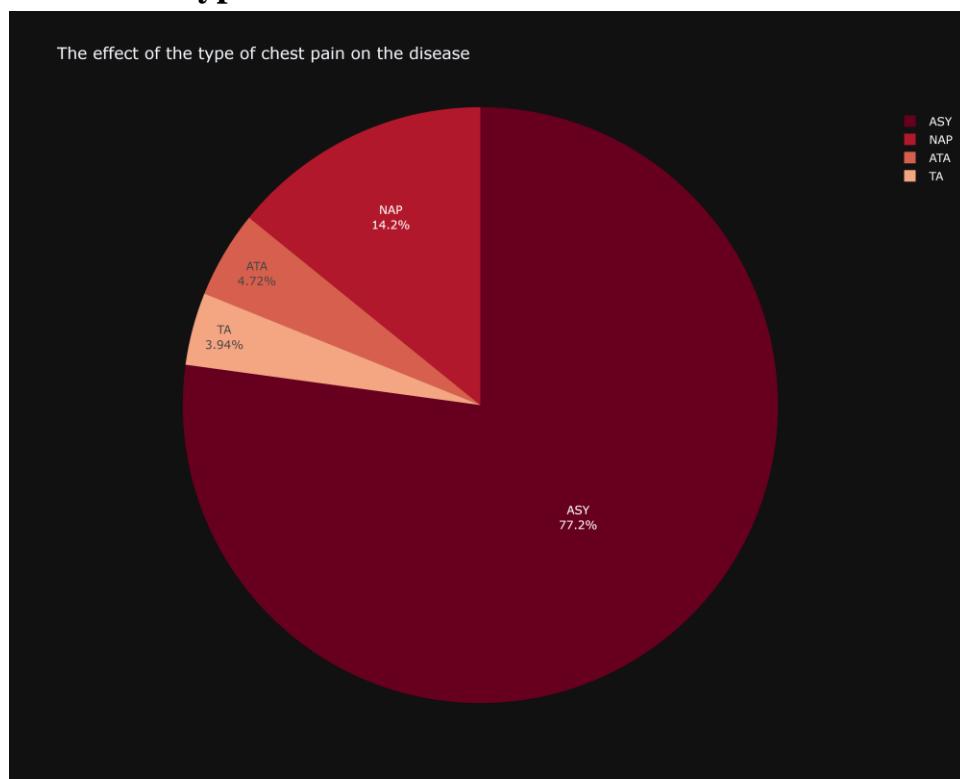
**График №28. Столбиковая диаграмма распределения переменных: 'Sex', 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST\_Slope', 'HeartDisease' по категориям**







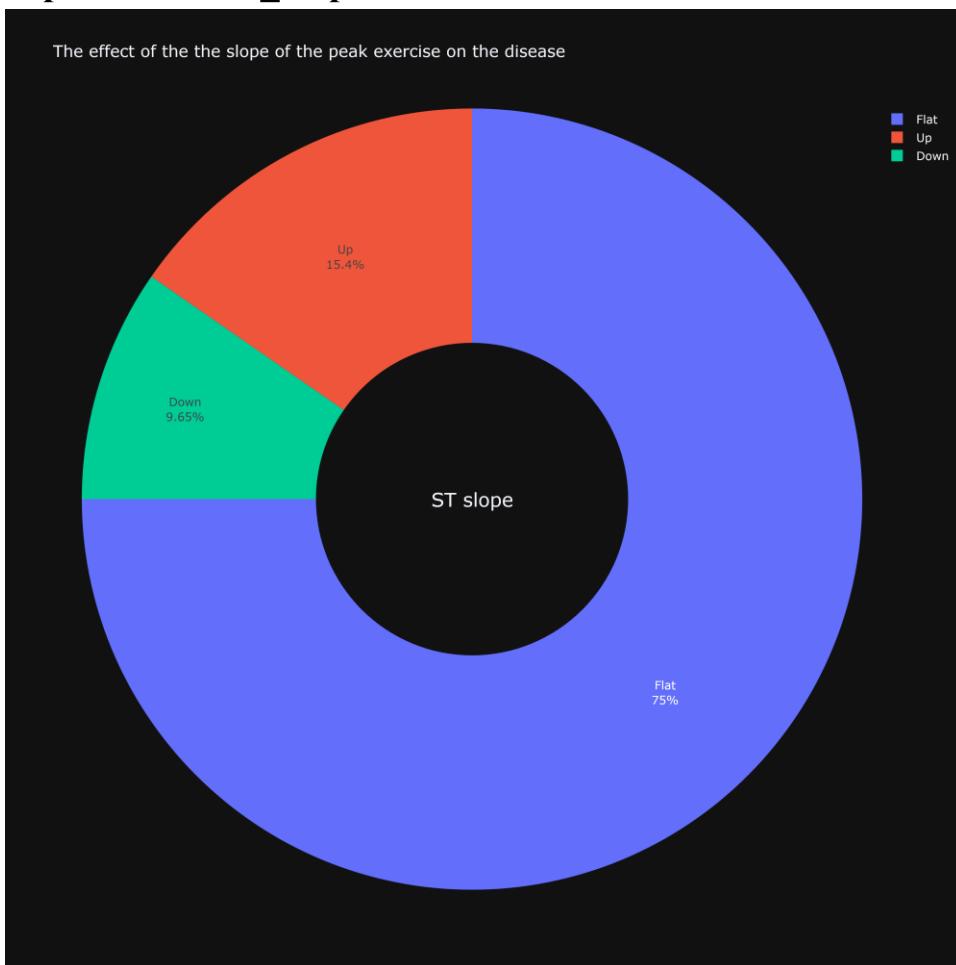
**График №29. Секторная диаграмма распределения переменной ChestPainType**







**График №30. Секторная диаграмма (2-го типа) распределения переменной ST\_Slope**







**График №31. Комбинированная диаграмма распределения переменной пол (Sex)**

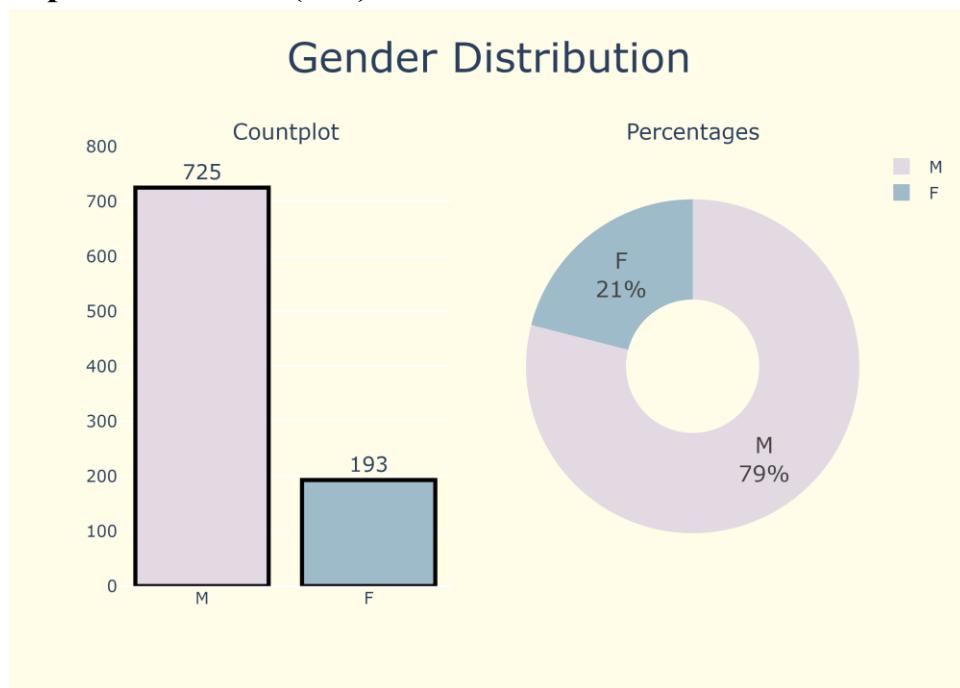






График №32. Комбинированная диаграмма распределения переменной HeartDisease

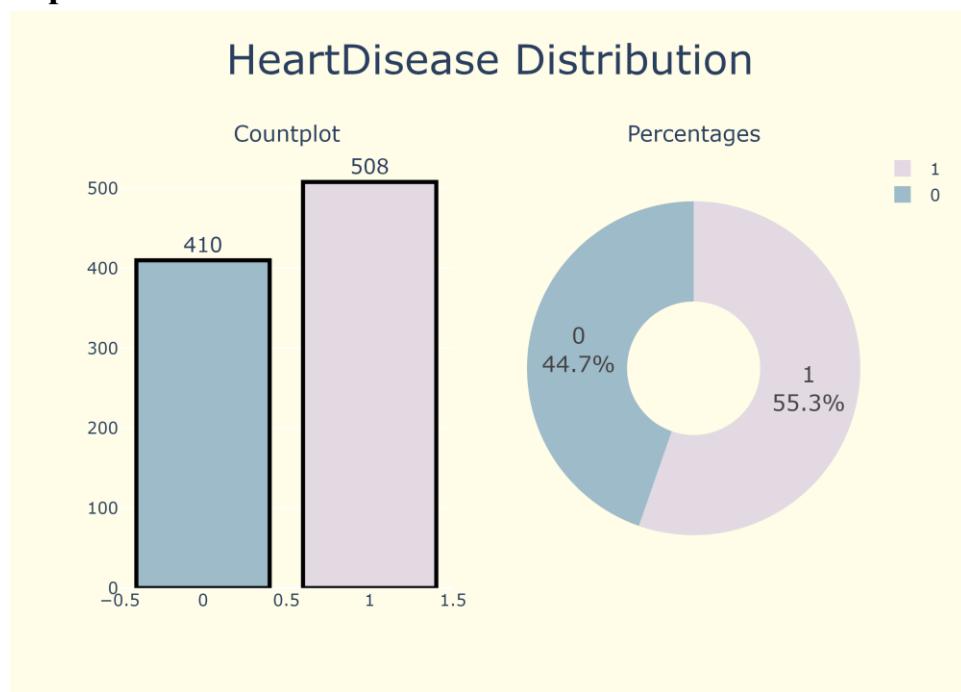
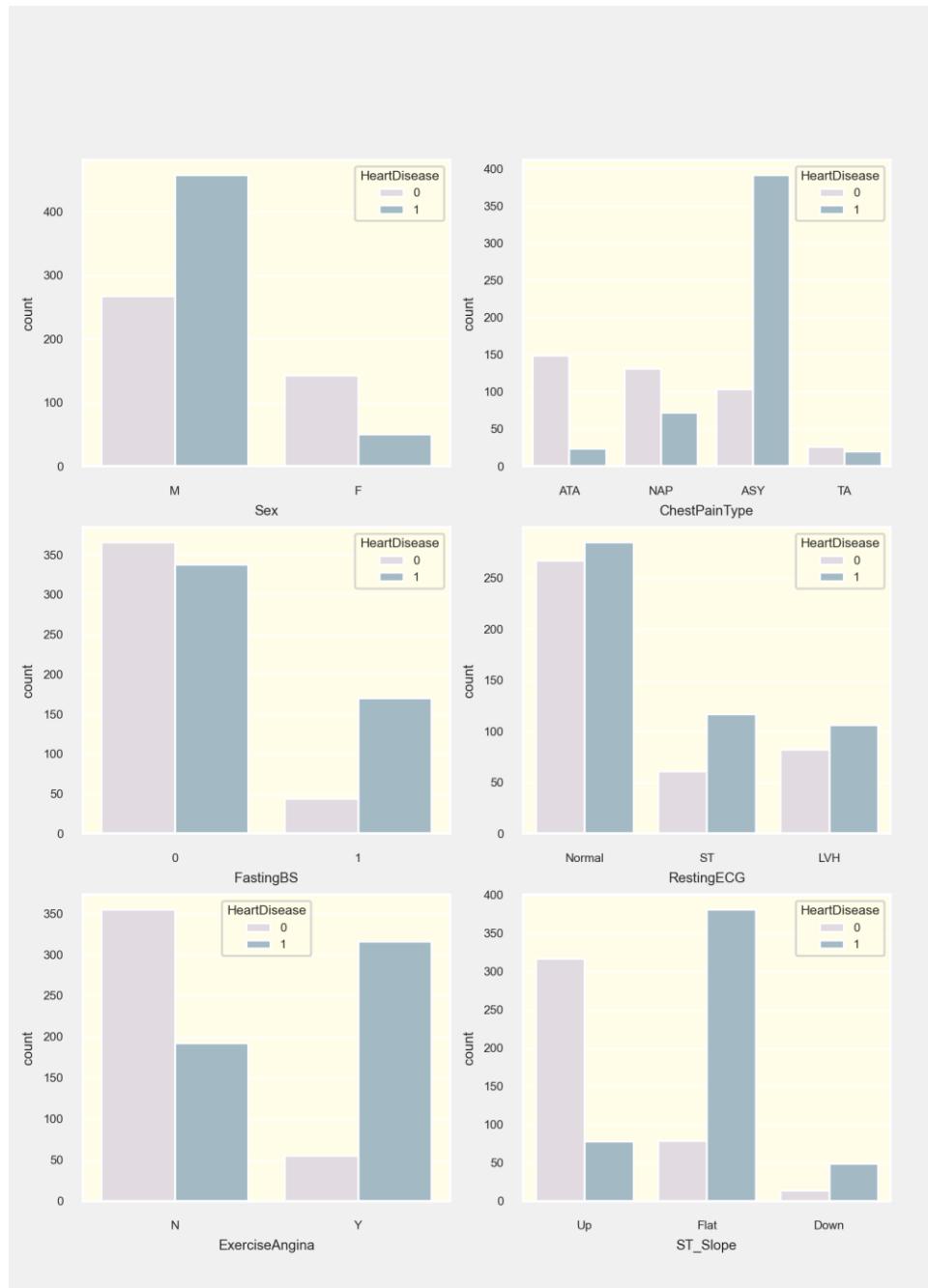






График №33. Столбиковая диаграмма распределения переменных: 'Sex', 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST\_Slope' по категориям и классу 'HeartDisease'







**График №34. Столбиковая диаграмма распределения переменных: 'ChestPainType', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST\_Slope' по категориям и классу 'Sex'**

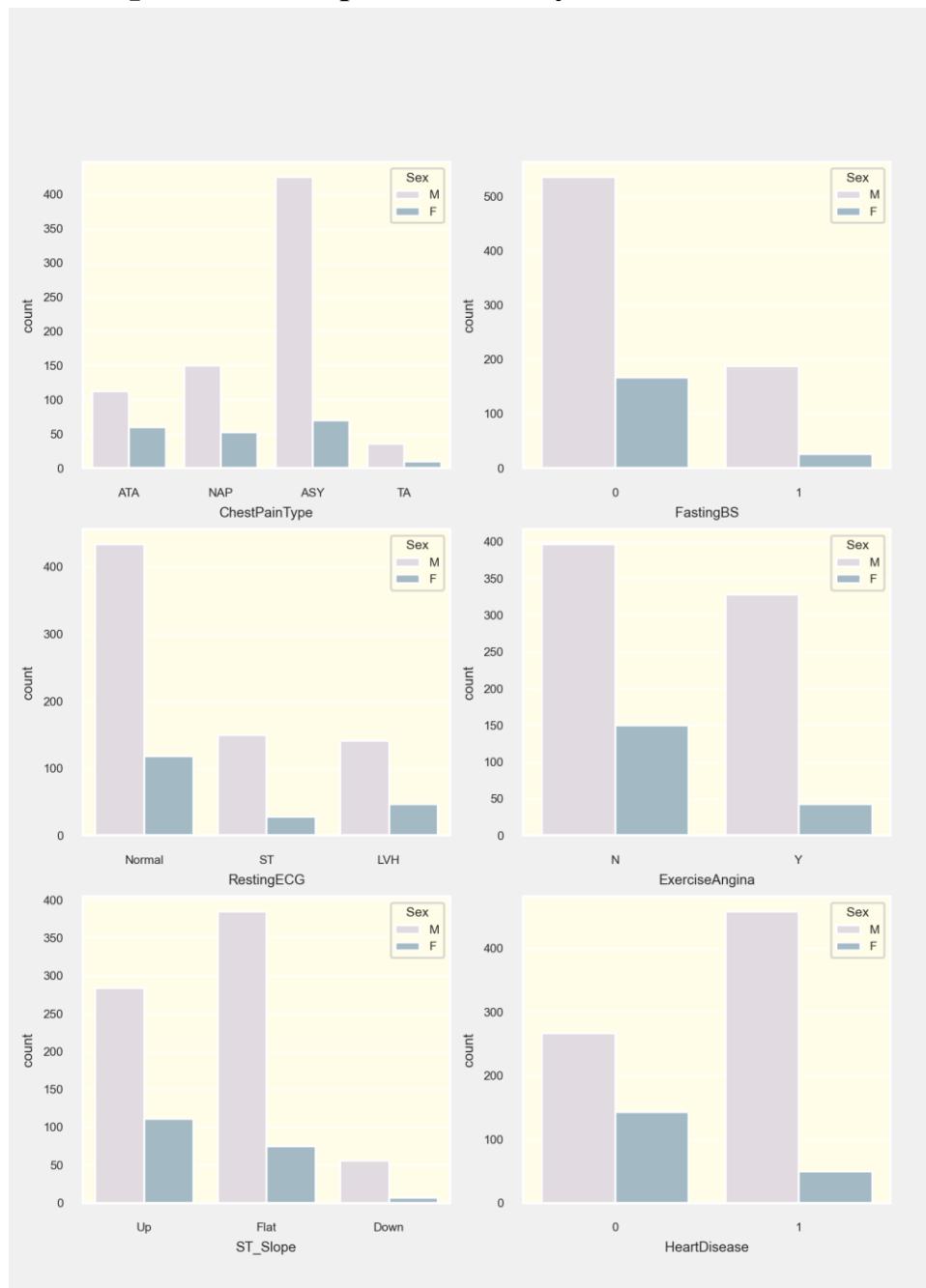
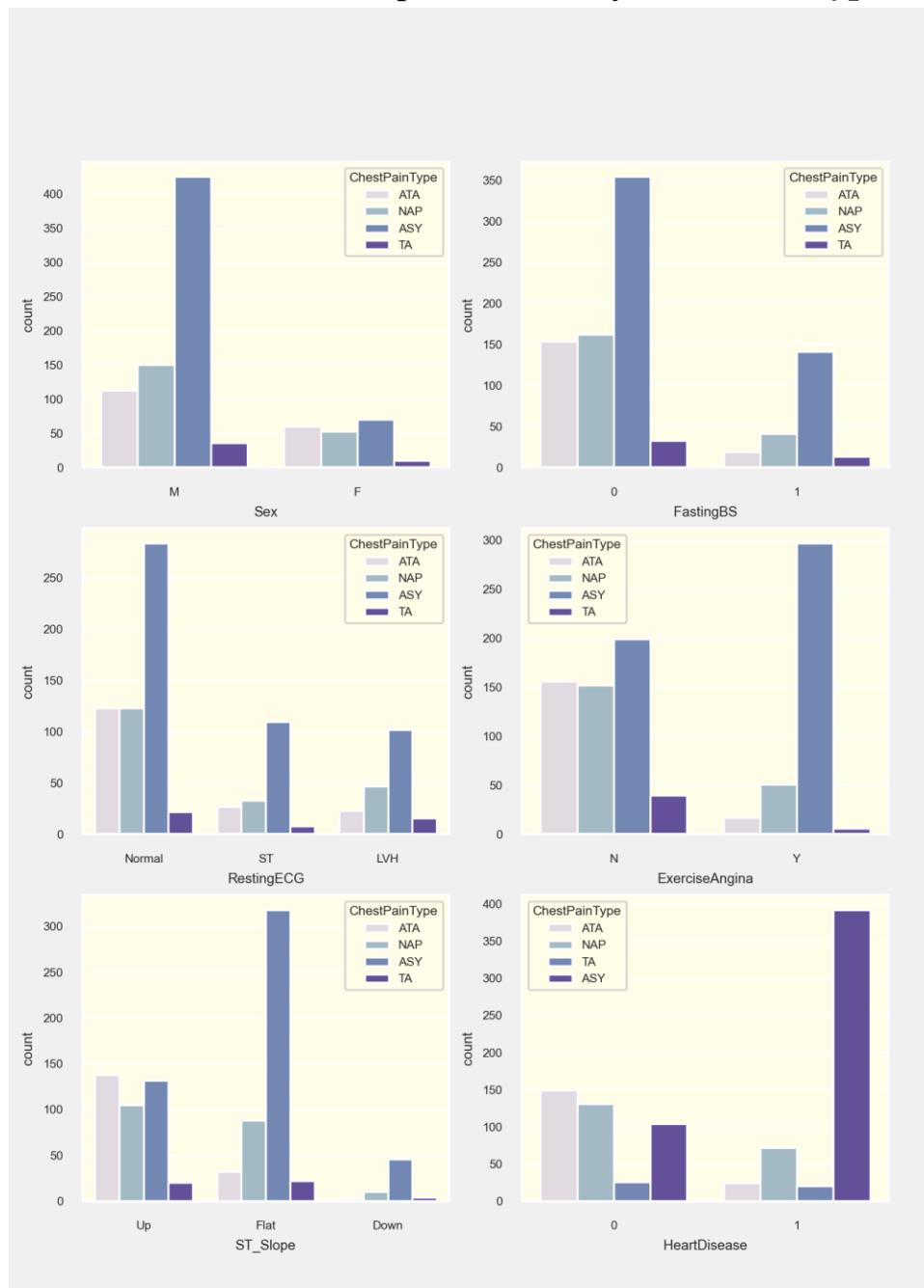






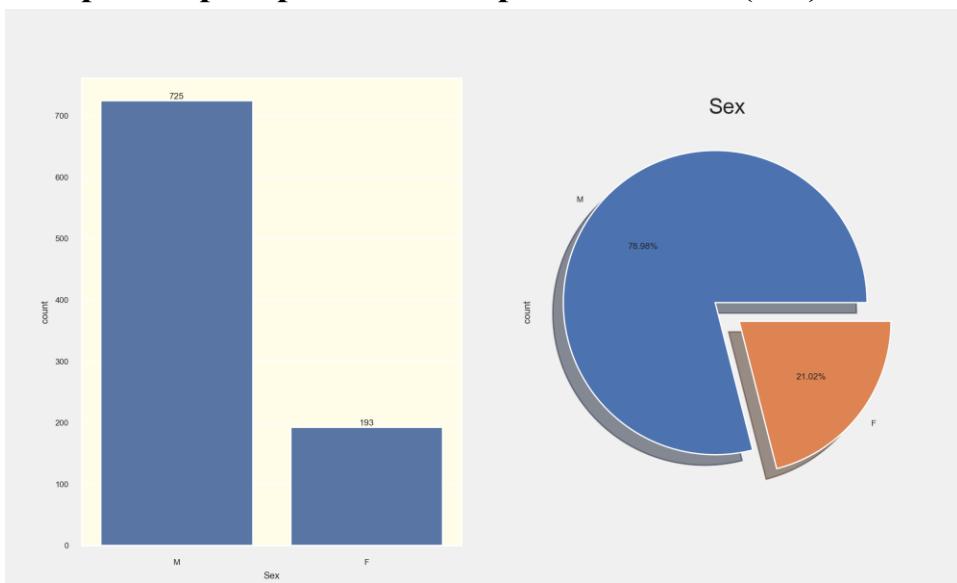
График №35. Столбиковая диаграмма распределения переменных: 'Sex', 'FastingBS', 'RestingECG', 'ExerciseAngina', 'ST\_Slope', 'HeartDisease' по категориям и классу 'ChestPainType'







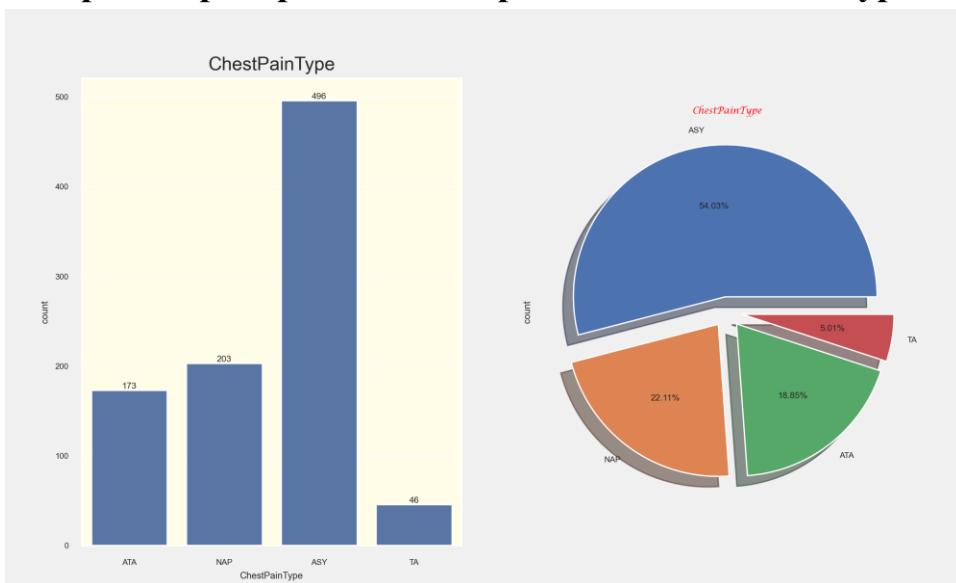
### График №36. Комбинированная (столбиковая и секторная) диаграмма распределения переменной пол (Sex)







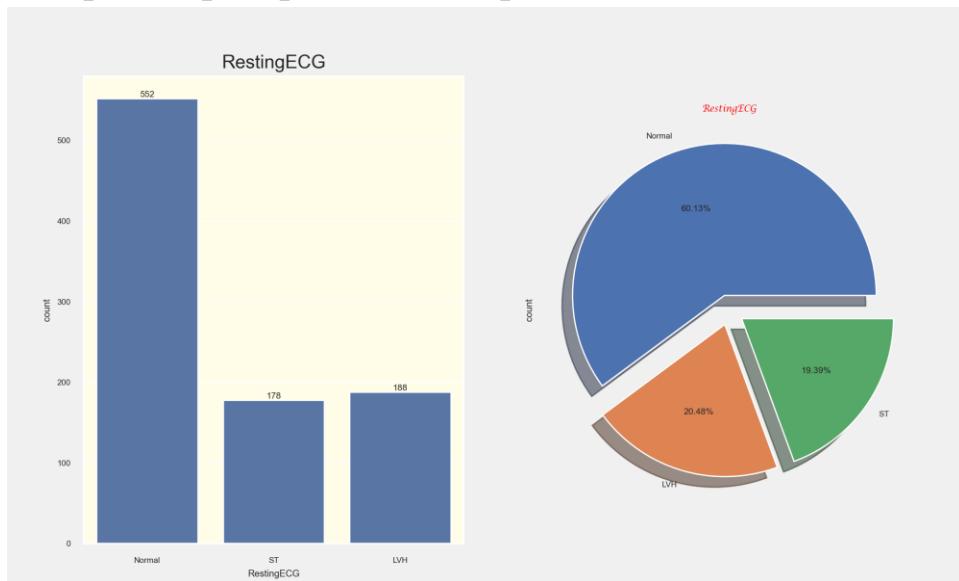
## График №37. Комбинированная (столбиковая и секторная) диаграмма распределения переменной ChestPainType







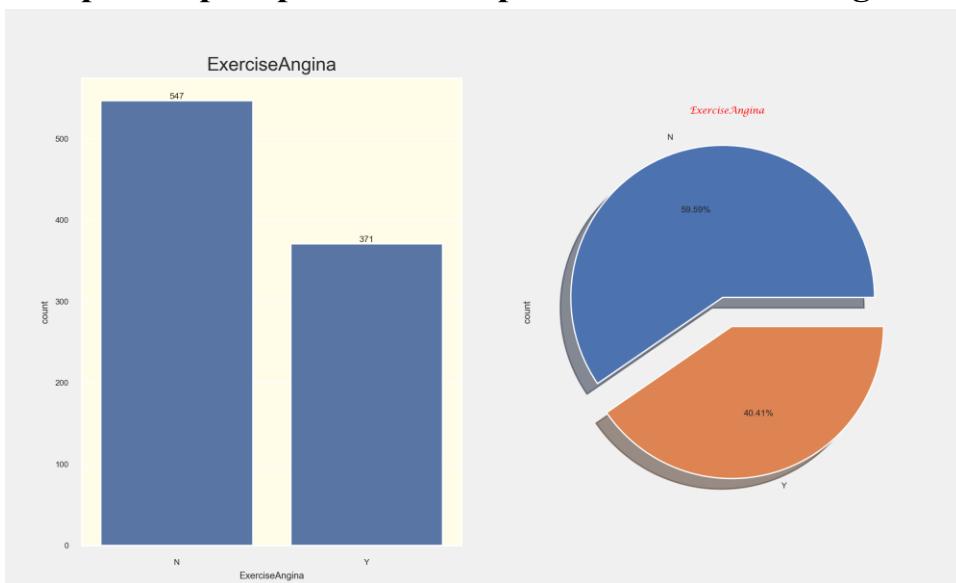
### График №38. Комбинированная (столбиковая и секторная) диаграмма распределения переменной RestingECG







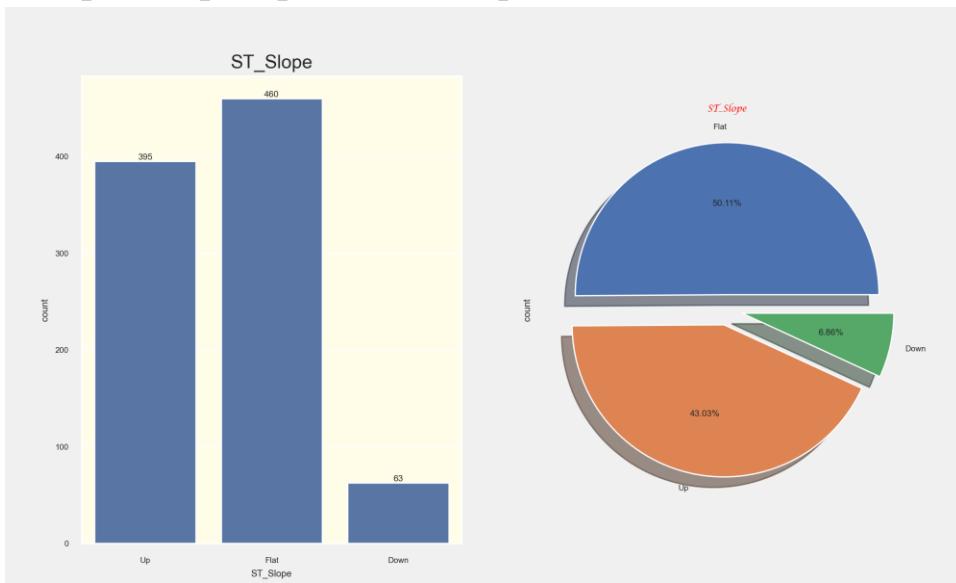
### График №39. Комбинированная (столбиковая и секторная) диаграмма распределения переменной ExerciseAngina







## График №40. Комбинированная (столбиковая и секторная) диаграмма распределения переменной ST\_Slope







## График №41. Комбинированная (столбиковая и секторная) диаграмма распределения переменной Cholesterol\_Category

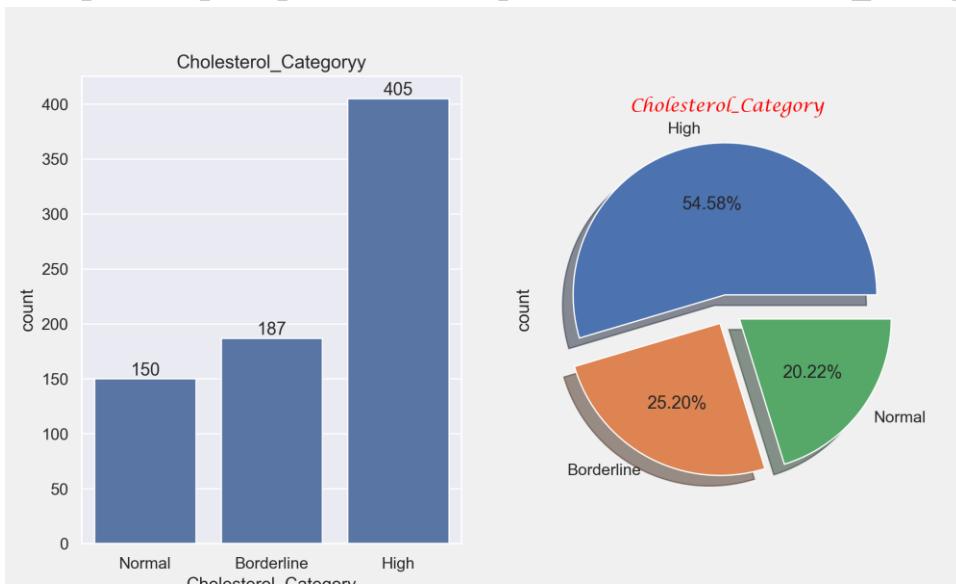
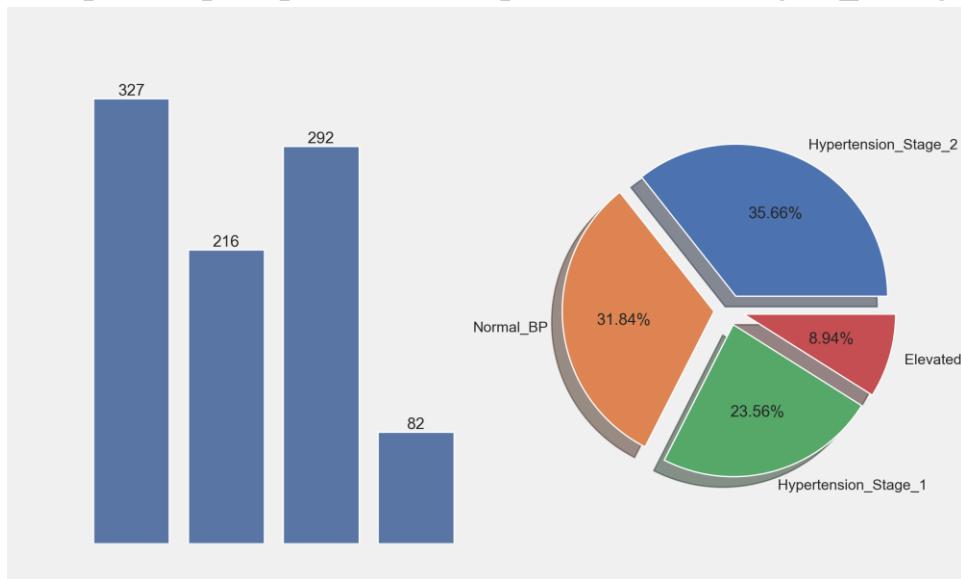






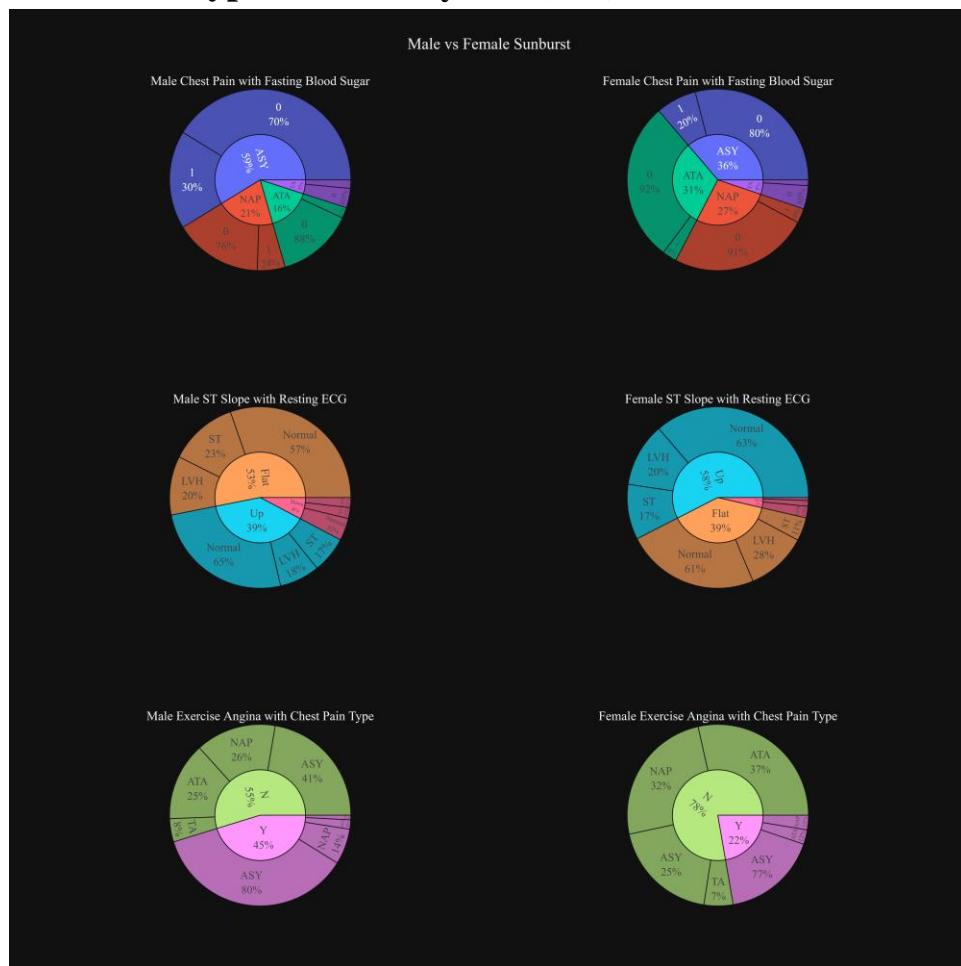
График №42. Комбинированная (столбиковая и секторная) диаграмма распределения переменной RestingBP\_Category







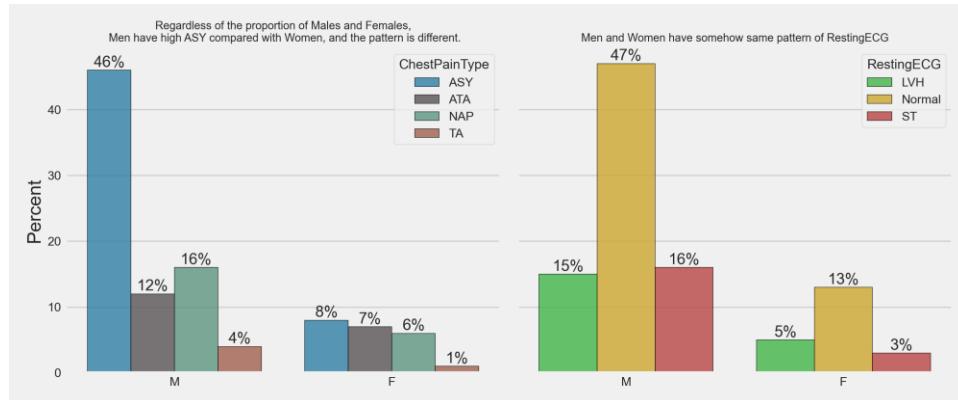
**График №43. Двойная секторная диаграмма (Sunburst, 6 субграфиков) распределения пар переменных: ['ChestPainType', 'FastingBS'], ['ST\_Slope', 'RestingECG'], ['ExerciseAngina', 'ChestPainType'] по классу пол (Sex)**







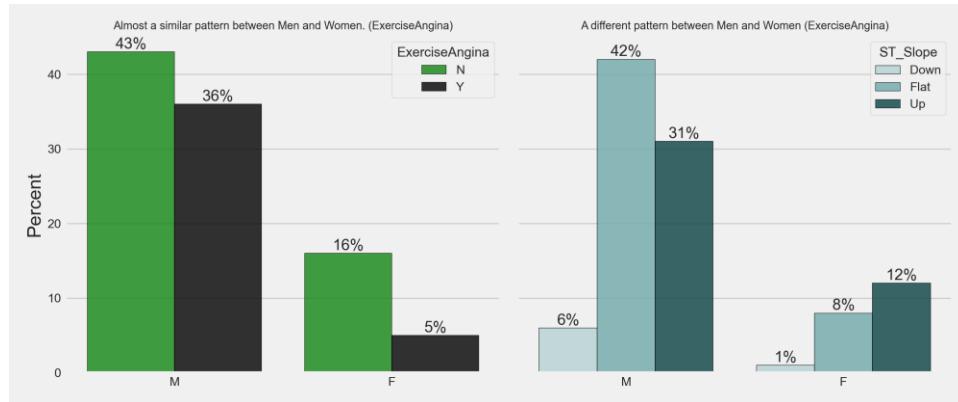
**График №44. Столбиковая диаграмма распределения переменных: RestingECG, ChestPainType (2 субграфика на одной панели) по классу пол (Sex)**







**График №45. Столбиковая диаграмма распределения переменных: ST\_Slope, ExerciseAngina (2 субграфика на одной панели) по классу пол (Sex)**





## Заключение по результатам EDA

EDA на наборе данных показал нам, как каждая переменная связана с переменной отклика и как мы можем сделать нашу модель эффективной, используя различные методы EDA. Визуализации данных поднимают наше понимание набора данных на более высокий уровень, позволяя нам делать выводы о направлении дальнейшего анализа.

Интеграция классификатора дерева принятия решений в наш анализ расширяет прогностические возможности нашей модели. В этом разделе мы не только изучили набор данных с помощью методов EDA, но и сделали еще один шаг вперед, внедрив модель машинного обучения. Такой целостный подход позволяет нам использовать сильные стороны как статистического анализа, так и прогнозного моделирования, способствуя более глубокому пониманию сложной динамики, связанной с прогнозированием ССЗ.