



# Применение методов машинного обучения в задаче анализа когорт пациентов с атеросклерозом артерий конечностей

---

М.В. Демченко AI и ML разработчик в DataArt, аспирант ВГУ

И.Л. Каширина, д.т.н., профессор ПММ ВГУ

# Цель исследования

---

Целью данной работы является поиск основных кластеров пациентов с диагностированным атеросклерозом артерий конечностей, с последующим анализом и выявлением основных закономерностей течения данного заболевания.

# Постановка задачи

---

В ходе данной работы требовалось:

- выделить основные кластеры, характеризующие состояние пациентов с диагностированным атеросклерозом
- проанализировать основные закономерности, обнаруженные в построенных кластерах

# Набор данных MIMIC-III

---



MIMIC-III – международная база данных с открытым доступом, содержащая информацию о поступлениях пациентов в отделение интенсивной терапии медицинского центра в Бостоне с 2001 по 2012 гг.

Достоинства:

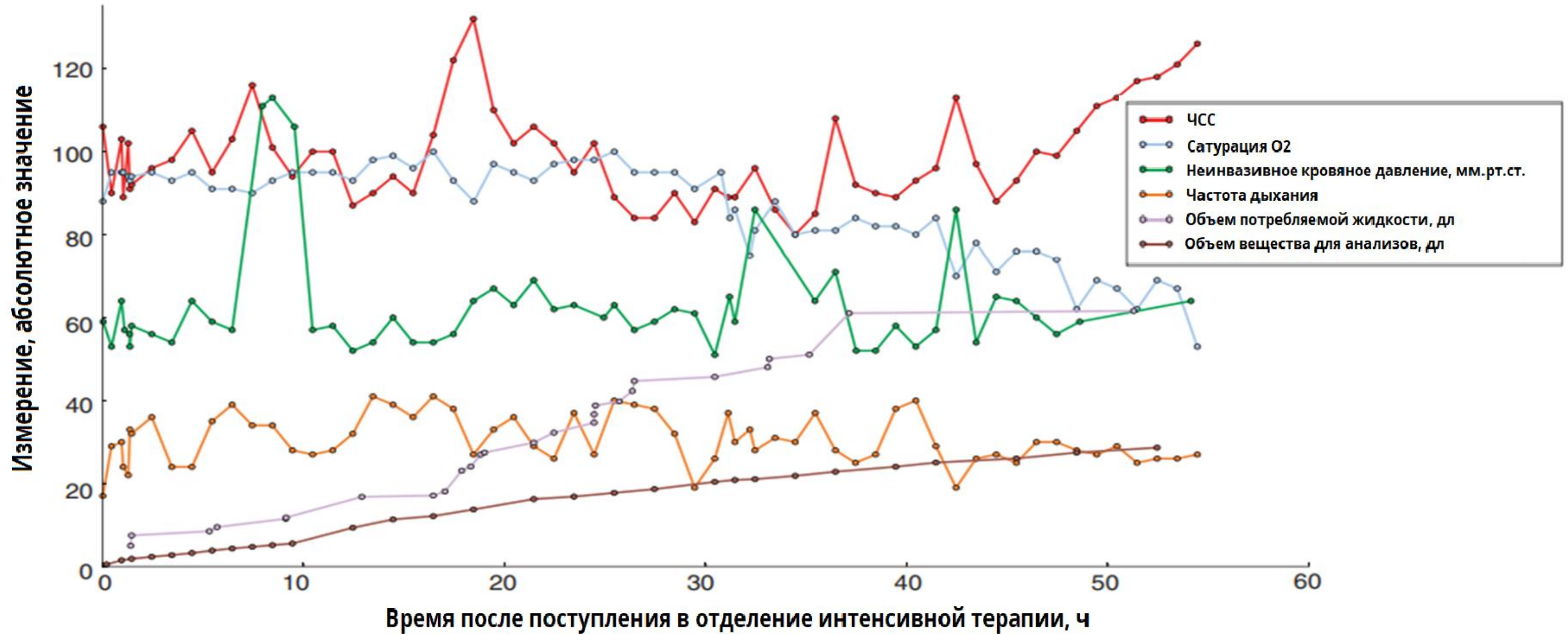
- Большой объем данных
- Содержит ряд основных показателей состояния организма, лабораторных измерений, медицинских предписаний и отчетов

# Основные признаки набора данных MIMIC-III



Категория признаков	Признаки
Гемодинамические	Arterial BP [Systolic] mmHg ( <i>систолическое артериальное давление</i> ), Arterial BP [Diastolic] mmHg ( <i>диастолическое артериальное давление</i> ), Heart Rate BPM ( <i>частота сердечных сокращений</i> ), Heart Rhythm ( <i>тип сердечного ритма</i> ), Respiratory Rate BPM ( <i>частота дыхания</i> ), SpO2 % и др.
Лабораторные	Glucose mg/dL ( <i>глюкоза</i> ), Calcium Total mg/dL ( <i>кальций</i> ), Creatinine mg/dL ( <i>креатинин</i> ), Hemoglobin g/dL ( <i>гемоглобин</i> ), Magnesium mg/dL ( <i>магний</i> ), Phosphate mg/dL ( <i>фосфаты</i> ), Platelet Count K/uL ( <i>тромбоциты</i> ), White Blood Cells K/uL ( <i>лейкоциты</i> ), Red Blood Cells m/uL ( <i>эритроциты</i> ), Sodium mEq/L ( <i>натрий</i> ), Potassium mEq/L ( <i>калий</i> ) и др.
Антропометрические	Height (cm) ( <i>Рост</i> ), Previous WeightF kg ( <i>вес</i> )
Социально-демографические	MARITAL_STATUS ( <i>семейное положение</i> ), GENDER ( <i>пол</i> ), AGE ( <i>возраст</i> )
Клинические	ADMISSION_TYPE ( <i>тип госпитализации</i> ), Ectopy Frequency ( <i>частота эктопии</i> ), Code Status ( <i>тип реабилитации</i> )
Указатели отклонения от нормы	Creatinine mg/dL Flag ( <i>Креатинин [Отклонение от нормы]</i> ), Glucose mg/dL Flag ( <i>глюкоза [Отклонение от нормы]</i> ), ABP Alarm [High] mmHg Flag ( <i>АД [Выше нормы]</i> ), ABP Alarm [Low] mmHg Flag ( <i>АД [Ниже нормы]</i> ), HR Alarm [High] BPM Flag ( <i>ЧСС[Выше нормы]</i> ), HR Alarm [Low] BPM Flag ( <i>ЧСС[ниже нормы]</i> )

# Показатели MIMIC-III



# Подготовка набора данных MIMIC III.

## Исходное представление данных



CHARTTIME	HADM_ID	ICUSTAY_ID	SUBJECT_ID	Arterial BP [Diastolic] mmHg	Arterial BP [Systolic] mmHg	Heart Rate BPM	Heart Rhythm	Respiratory Rate BPM	SpO2 %
2165-05-30 08:20:00	134369	225073	124	45	118.0	100.0	Sinus Tachy	14.0	93.0
2165-05-30 09:00:00	134369	225073	124	46	115.0	100.0	Sinus Tachy	19.0	96.0
2165-05-30 10:00:00	134369	225073	124	51	119.0	101.0	Sinus Tachy	14.0	98.0
2165-05-30 10:30:00	134369	225073	124	51	117.0	101.0	Sinus Tachy	22.0	98.0
2165-05-30 11:00:00	134369	225073	124	46	116.0	101.0	Sinus Tachy	20.0	97.0
2165-05-30 12:00:00	134369	225073	124	44	106.0	97.0	Normal Sinus	18.0	96.0
2165-05-30 13:00:00	134369	225073	124	47	131.0	99.0	Normal Sinus	16.0	96.0

# Подготовка набора данных MIMIC III. Нормализованное представление данных



ICUSTAY_ID	SUBJECT_ID	HADM_ID	CHARTTIME	Arterial BP [Diastolic] mmHg	Arterial BP [Systolic] mmHg	Heart Rate BPM	Previous WeightF kg	Respiratory Rate BPM	SpO2 %	HR Alarm [High] BPM Flag	HR Alarm [Low] BPM Flag	ABP Alarm [High] mmHg Flag	ABP Alarm [Low] mmHg Flag
249669	6597	112631	2100-09-18 16:00:00	26.000000	54.000000	107.500000	78.0	19.000000	96.00	0	0	0	2
249669	6597	112631	2100-09-18 18:00:00	32.714286	79.142857	97.642857	78.0	16.214286	97.00	0	0	0	10
249669	6597	112631	2100-09-18 20:00:00	29.666667	93.666667	95.666667	78.0	14.666667	97.00	0	0	0	2
249669	6597	112631	2100-09-19 00:00:00	30.500000	110.000000	104.500000	78.0	23.500000	97.50	0	0	0	0
249669	6597	112631	2100-09-19 02:00:00	34.000000	109.500000	97.000000	78.0	18.000000	99.00	0	0	0	0
249669	6597	112631	2100-09-19 04:00:00	34.500000	117.500000	92.000000	78.0	17.500000	99.50	0	0	0	0



# Набор данных MIMIC-III

Полученный набор данных  
содержит 3443 записи о 114  
пациентах в рамках 118  
госпитализаций по 32  
показателям



Features	Count	Rate, %
GENDER_M	1940	56.35
Ectopy Frequency _Frequent	150	4.36
Ectopy Frequency _None	2505	72.76
Ectopy Frequency _Occasional	453	13.16
Heart Rhythm _1st Deg AV Block	108	3.14
Heart Rhythm _2nd AVB Mobitz 2	1	0.03
Heart Rhythm _2nd AVB/Mobitz I	22	0.64
	mean	std
Arterial BP [Diastolic] mmHg	54.81	17.39
Arterial BP [Systolic] mmHg	121.24	24.16
Heart Rate BPM	85.32	15.93
SpO2 %	97.29	3.83
Calcium Total mg/dL	8.25	0.73
Chloride mEq/L	104.66	5.59
Creatinine mg/dL	1.90	1.81

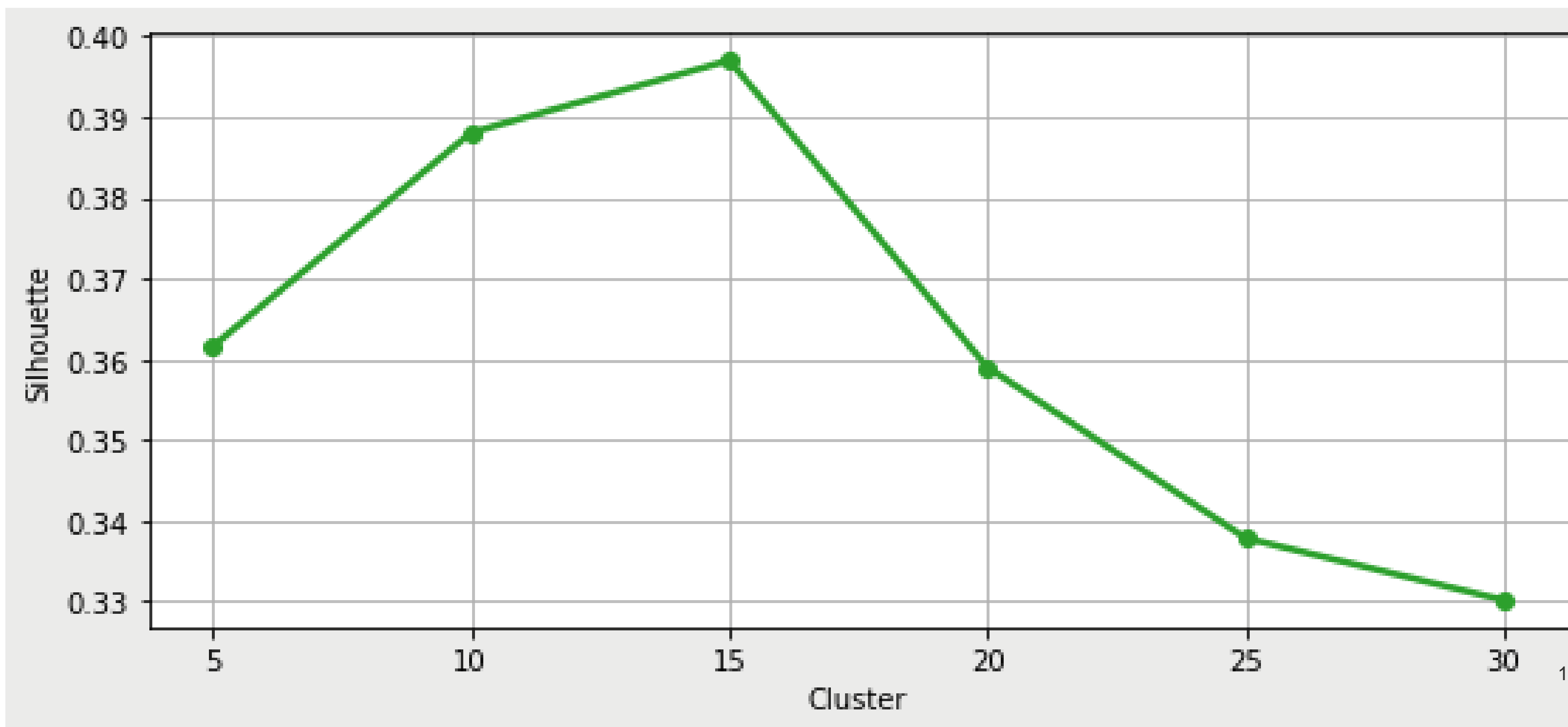
## Понижение размерности с помощью метода t-SNE

- Алгоритм нелинейного понижения размерности
- Алгоритм основан на сохранении локальной структуры данных с использованием расстояния Кульбака-Лейблера
- Один из наиболее эффективных и востребованных методов

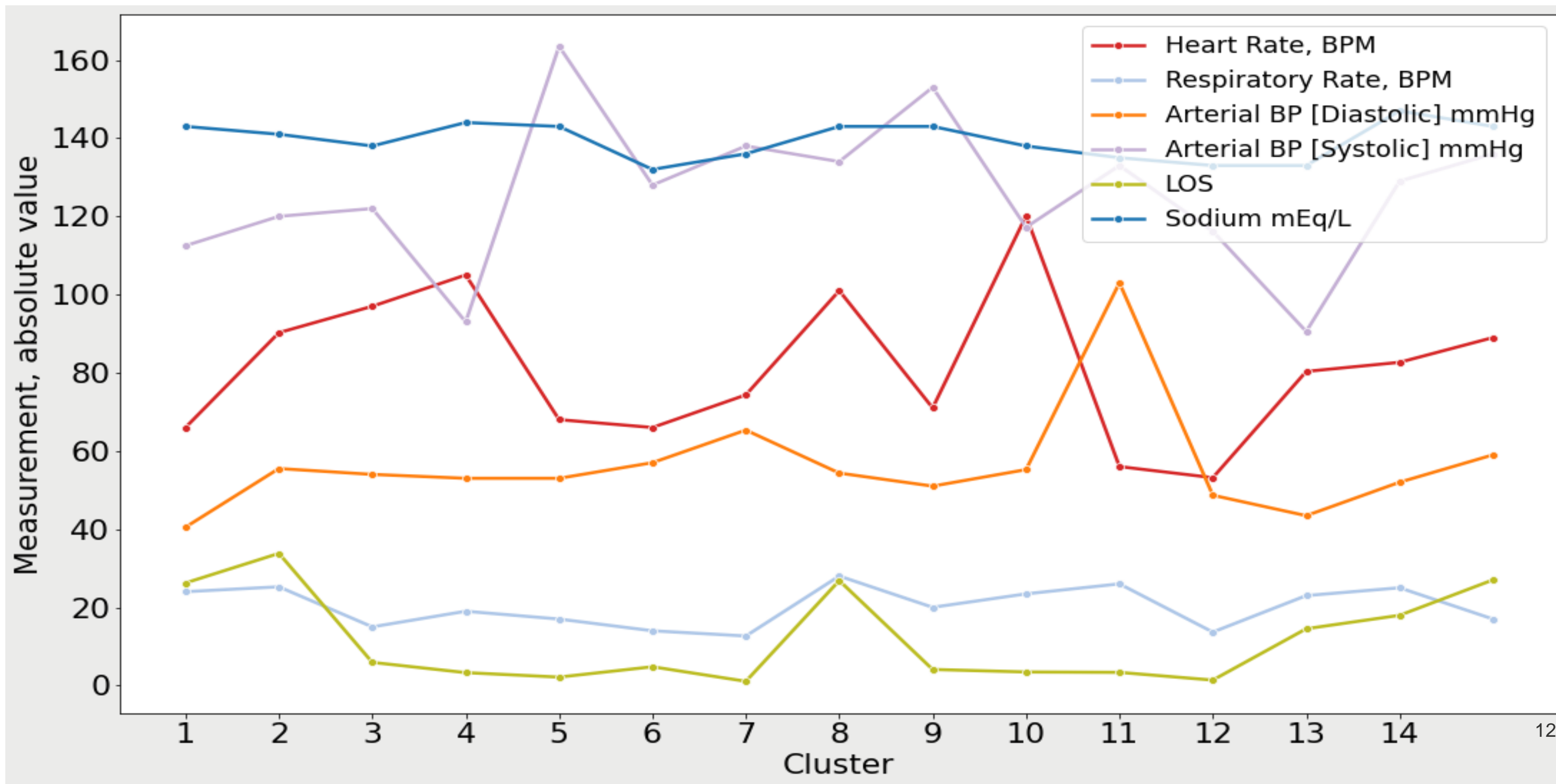
## Кластеризация с помощью метода k-medoids

- Алгоритм формирования кластеров вокруг медоидов – центральных точек, являющихся реальными объектами данных
- Выбор медоидов производится алгоритмически с помощью метода PAM (Partitioning around medoids)

# Выбор оптимального числа кластеров с помощью коэффициента силуэта

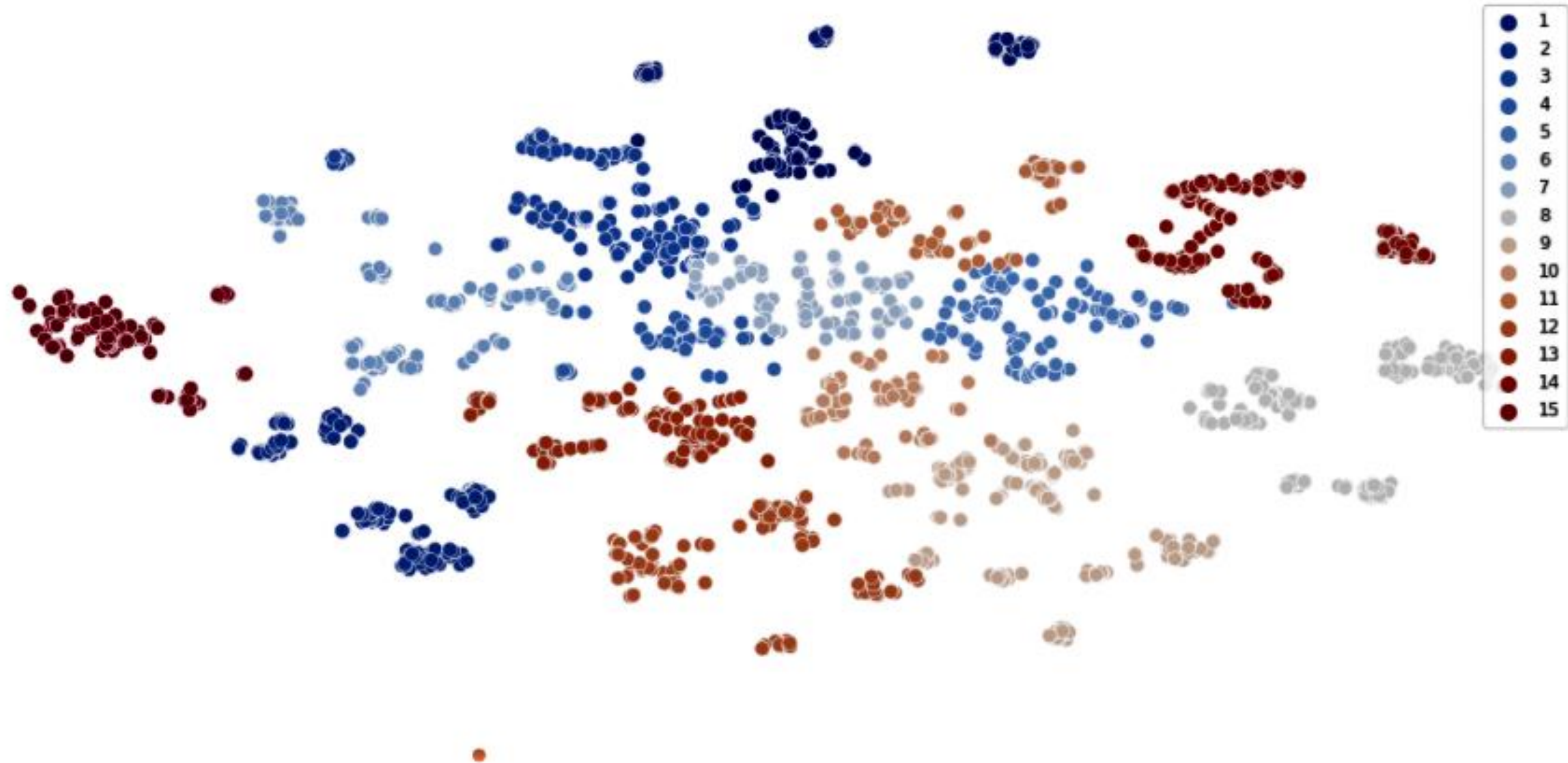


# Результаты. График медоидов



# Результаты. Визуализация кластеров

---



# Результаты. Анализ полученных кластеров



Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Expired_Rate	0	0	5.263	7.692	10	7.692	15.79	16.67	15.79	26.92	25	31.43	61.54	100	100
Creatinine [Abnormal], %	9.4	4.73	12.31	9.08	15.26	29.25	9.21	12.92	23.96	9.61	5.76	18.98	9.38	23.44	35.47
Glucose mg/dL [Abnormal]	19.88	27.88	26.02	27.55	21.95	18.59	17.56	23.13	27.36	16.02	19.62	21.83	17.22	19.14	26.35
HR Alarm [High]	0.14	0.72	0.17	3.03	1.23	0.52	1.85	0.36	0.81	2.42	0.54	0.28	4.3	0.25	0.17
HR Alarm [Low]	4.7	1.04	14.61	12.58	8.52	14.16	3.38	2.81	12.97	4.55	13.41	6.57	4.42	0	32.85
ABP Alarm [High]	14.17	10.42	11.46	5.73	8.98	8.68	1.69	3.65	8.97	6.52	14.58	9.74	7.47	3.41	23.31
ABP Alarm [Low]	2.6	3.93	7.5	12.26	5.1	13.23	4.69	3.43	9.17	5.34	4.95	2.76	6.66	0.21	9.71
Bicarbonate [Abnormal]	11.14	11.06	11.16	2.87	8.01	13.29	10.91	1.83	18.49	10.29	14.49	9.93	9.27	10.09	2.96
Calcium [Abnormal]	17.14	26.44	28.07	27.39	18.84	13.05	15.43	16.13	25.84	13.21	18.9	16.27	11.7	6.43	11.74
Chloride [Abnormal]	9.11	3.53	4.98	4.94	5.56	6.53	6.71	12.08	9.22	5.4	11.61	9.56	4.53	10.83	3.8
Hemoglobin [Abnormal]	29.36	37.98	31.05	32.8	26.19	30.77	25.25	26.07	32.17	18.77	27.9	24.4	21.27	23.3	35.47
Magnesium [Abnormal]	1.81	0.08	1.11	3.5	3.42	0.87	2.02	0	1.17	1.74	0.45	0.87	1.55	0.25	6.25
Phosphate [Abnormal]	7.09	17.95	18.95	12.9	9.65	22.79	9.87	8.29	14.84	9.33	6.75	14.34	12.55	15.02	25.17
Potassium [Abnormal]	4.41	0.4	3.66	7.48	1.79	8.33	2.18	0.62	3.04	0.17	3.51	3.12	2.65	2.7	5.66
Sodium [Abnormal]	1.74	0.16	5.96	7.64	1.28	10.37	6.38	5.26	5.17	3.32	6.75	6.62	4.71	9.8	2.28
White Blood Cells [Abnormal]	18.66	23.56	21.76	19.9	20.01	27.68	12.65	17.34	21.23	15.4	23.04	16.5	14.86	21.2	9.88
Resp Alarm [High]	0	0.32	0.3	1.27	0.1	0.47	0.22	2.01	0.2	0.17	0	0.14	0.26	0.04	0.08
Resp Alarm [Low]	5.5	1.44	14.74	16.4	9.49	15.21	4.85	2.72	14.29	8.83	14.04	6.07	6.99	0.36	33.02
SpO2 Alarm [Low]	5.13	2.24	14.74	16.88	10.21	14.34	5.23	4.46	13.83	7.03	14.4	6.66	11.04	0.64	33.28
Anion Gap [Abnormal]	0.29	1.52	5.54	0.96	0.51	11.36	1.58	1.38	2.58	1.46	2.25	1.24	4.56	0.67	4.81
Hematocrit [Abnormal]	29.86	37.98	31.05	32.8	28.23	30.77	24.97	25.58	32.93	18.1	28.08	27.3	21.27	23.44	35.47
Platelet [Abnormal]	3.11	12.02	3.92	10.19	13.68	8.16	4.91	2.94	13.83	11.35	11.79	14.8	11.66	8.91	6.25 <sup>14</sup>
Red Blood Cells [Abnormal]	30.73	37.98	31.26	33.6	28.23	30.77	24.97	24.96	32.93	18.94	28.08	27.48	21.27	23.47	35.47
Urea Nitrogen [Abnormal]	11.28	9.29	16.23	9.08	15.31	29.25	14.01	20.77	28.67	13.66	16.38	21.6	7.25	22.76	35.47

Спасибо за внимание!