	 75.0 55.0 65.0 50.0 65.0 	_	nosphokina 5 78 1 1	se diabetes	_	n high_blood_pres	1 265000.00 0 263358.03 0 162000.00 0 210000.00 0 327000.00 1 204000.00	1.90 1.10 1.30 1.90 2.70 2.10	erum_sodiu 13 13 12 13 11
	7 60.0 8 65.0 9 80.0 10 75.0 11 62.0 12 45.0 13 50.0	1 1 0 1 1 0 1 1 1 1 1 1 1 1 1	3 1 1 2 9	46 0 15 1 57 0 23 0 81 0 81 0 68 0	1 6 6 3 3 2 3 3 3	0 5 5 8 5 0	0 127000.00 0 454000.00 0 263358.03 1 388000.00 1 368000.00 0 136000.00 1 276000.00 1 427000.00	1.20 1.10 1.50 9.40 4.00 0.90 1.10 1.10	13 13 13 13 14 13 13
	16 87.0 17 45.0 18 70.0 19 48.0 20 65.0 21 65.0 22 68.0	1 1 0 1 1 1 1 1 1 0	1. 5 1 5	79 0 49 0 82 0 25 0 82 1 52 0 28 1 20 0	5 3 1 2 5 2 3 3 6	8 4 5 5 5 0	0 47000.00 0 262000.00 0 166000.00 1 237000.00 0 87000.00 1 276000.00 1 297000.00 1 289000.00 0 368000.00	1.30 0.90 0.80 1.00 1.90 1.30 1.60 0.90	13 14 12 14 13 13 14
	 75.0 80.0 95.0 70.0 58.0 82.0 94.0 85.0 	0 0 1 0 1 0 0	5 1 1 1	82 1 48 1 12 0 22 1 60 0 70 1 82 1 23 0	3 3 4 4 3 3 3 4	0 8 0 5 8 0 8	1 263358.03 0 149000.00 1 196000.00 1 284000.00 0 153000.00 0 200000.00 1 263358.03 0 360000.00	1.83 1.90 1.00 1.30 5.80 1.20 1.83 3.00	13 14 13 13 13 13
	33 50.0 34 65.0 35 69.0 36 90.0 37 82.0 38 60.0 39 60.0	1 1 0 0 1 1 1 0 0 0 0 0 0 0	1 5 8 26 2	49 1 59 1 94 1 82 1 60 1 55 1 56 1 35 1 82 0	3 3 5 3 5 5 3 3 3 2	0 0 5 0 0 0	1 319000.00 0 302000.00 1 188000.00 0 228000.00 0 226000.00 1 321000.00 0 305000.00 0 329000.00 1 263358.03	1.00 1.20 1.00 3.50 1.00 1.00 2.30 3.00 1.83	12 13 14 13 14 13 14 13
	42 70.0 43 72.0 44 60.0 45 50.0 46 51.0 47 60.0 48 80.0	0 0 0 1 0 0 0	5 1 5 5 13 5	24 1 71 1 27 1 88 1 82 1 80 0 82 1 53 0 29 0	3 4 5 6 3 2 3 2	5 0 0 8 5 8	1 153000.00 1 185000.00 1 218000.00 0 194000.00 0 310000.00 1 271000.00 1 451000.00 1 140000.00 0 395000.00	1.20 1.20 1.00 1.10 1.90 0.90 0.60 4.40 1.00	13 13 14 13 13 13 13
2 2 2 2 2 2 2	 49 53.0 50 50.0 51 55.0 52 50.0 53 70.0 54 53.0 55 52.0 	 0 0 0 0 0 1 1	2 25 5 2 4	 07 1		 0 0 5 5 5 0		1.20 0.50 0.80 1.00 1.20 1.00 1.70	13 14 13 13 13 14 14
2 2 2 2 2 2 2 2 2	57 58.0 58 45.0 59 53.0 60 55.0 61 62.0 62 65.0 63 68.0 64 61.0	0 1 0 0 1 1 1	1 6 2 1 5	32 1 66 1 56 0 66 0 55 0 58 1 57 1 82 1	3 2 5 4 4 2 6 3	8 5 0 0 0 5 0	1 253000.00 0 233000.00 0 308000.00 0 203000.00 0 283000.00 0 198000.00 0 208000.00 0 147000.00 0 362000.00	1.00 0.80 0.70 1.00 0.70 1.40 1.00 1.20 0.90	13 13 13 13 14 14 14
2 2 2 2 2 2 2 2 2	66 55.0 67 56.0 68 45.0 69 40.0 70 44.0 71 51.0 72 67.0 73 42.0	1 0 1 0 0 0 0 0	11 1 5 5 5 5 2		2 3 3 3 3 4 3 4	0 8 8 8 5 0 0	0 263358.03 0 133000.00 0 302000.00 0 222000.00 1 263358.03 0 221000.00 0 215000.00 0 189000.00 0 150000.00	1.83 1.70 0.90 1.00 1.60 0.90 1.20 0.70	1: 14 14 15 11 11 14 14
2 2 2 2 2 2 2 2 2	75 45.0 76 70.0 77 70.0 78 50.0 79 55.0 80 70.0 81 70.0 82 42.0	0 0 0 1 0 0 0	5 6 5 10 26	82 0 18 0 82 1 51 1 84 1 95 1 82 0 64 0	3 3 3 3 4 4 4 3	8 5 8 0 8 0 0	1 422000.00 0 327000.00 0 25100.00 0 232000.00 0 451000.00 0 241000.00 0 51000.00 0 215000.00 0 263358.03	0.80 1.10 1.10 0.70 1.30 1.00 2.70 3.80 1.10	1; 14 14 1; 1; 1; 1; 1;
2 2 2 2 2 2 2 2	84 50.0 85 55.0 86 60.0 87 45.0 88 65.0 89 90.0 90 45.0 91 60.0	1 1 0 0 0 0 1 0	1 2 5 8 3 6	54 0 70 1 53 0 82 1 92 1 37 0 15 1 20 0	4 4 3 5 3 3 5	0 0 5 5 5 8 8 5	0 279000.00 0 336000.00 0 279000.00 0 543000.00 0 263358.03 0 390000.00 0 222000.00 0 133000.00	0.80 1.20 1.70 1.00 1.10 0.90 0.80 1.40	1. 1. 1. 1. 1. 1.
2 2 2 2 2 2 2	93 63.0 94 62.0 95 55.0 96 45.0 97 45.0 98 50.0 19 rows × 13 colu		1 18 20 24	60 1	3 3 3 6 3 4	5 8 8 0 8	0 382000.00 0 179000.00 1 155000.00 0 270000.00 0 742000.00 0 140000.00 0 395000.00	1.00 0.90 1.10 1.20 0.80 1.40 1.60	14 13 14 13 14 14
	creatinine pho diabetes: if the ejection fraction platelets: plate sex: woman or serum creatini	e patient (years ease of red bloosessure: if the passphokinase (CF e patient has dient percentage elets in the blooses man (binary)	od cells or atient has l PK): level o abetes (bo of blood l od (kilopla	hypertension (of the CPK enzy colean) eaving the heatelets/mL)	(boolean) yme in the bloa art at each con od (mg/dL)	od (mcg/L) traction (percenta	age)		
3]: a a c c d e h		e patient smoke p period (days) the patient de sphokinase	float6 int6 int6 int6 int6 int6 int6 int6	boolean) ring the follow 64 64 64 64 64 64		oolean)			
s s s s t D d d	erum_creatini erum_sodium ex moking ime EATH_EVENT type: object df.describe() age ount 299.000000 nean 60.833893	anaemia cr	float6 int6 int6 int6 int6	54 54 54 54 54 54 64 hosphokinase	diabetes ejo 299.000000 0.418060	ection_fraction hig 299.000000 38.083612	gh_blood_pressure 299.000000 0.351171	platelets se 299.000000 263358.029264	rum_creatir 299.00 1.39
5]: (std 11.894809 min 40.000000 25% 51.000000 50% 60.000000 75% 70.000000 max 95.000000 df.isnull().st	0.496107 0.000000 0.000000 0.000000 1.000000		581.839465 970.287881 23.000000 116.500000 250.000000 582.000000 7861.000000	0.418060 0.494067 0.000000 0.000000 0.000000 1.000000 1.000000	38.083612 11.834841 14.000000 30.000000 38.000000 45.000000 80.000000	0.478136 0.000000 0.000000 0.000000 1.000000	263358.029264 97804.236869 25100.000000 212500.000000 262000.000000 303500.000000 850000.000000	1.39 1.03 0.50 0.90 1.10 1.40 9.40
a c d d e h p p s s s s t D d		ion ssure ne aemia', 'dia				'sex' '	','DEATH_EVENT	']	
Be co ch	for ax, title snum_xar = [i : efore making any anclude that any aracteristics. Cat fig, axs = pla ax_title_pairs for ax, title sns.count ax.set_title	<pre>for i in df. visualisations, actions in order var are catergo t.subplots(n) s = zip(axs. in ax_title plot(x=title, f)</pre>	it's good to record to change orical variations rows=2, flat, capairs: , data=d	if i not in to check the co e the values in ables and num ncols=3, fi t_var)	olumn types, b n dataset are n n_var are nume .gsize=(14,	asic statistics and ot necessary. Varia rical variables.	number of NULLs	. Based on these s	•
tinos	ax.set_xla	abel('')	∠e	175 - 150 - 125 - 100 -	dia	abetes	200 - 175 - 150 - 125 - til 100 -	high_blood_pi	ressure
	60 - 40 - 20 - 0 0	sex	1		o sn	noking	75 - 50 - 25 - 0 200 - 175 - 150 -	DEATH_EV	ENT
W				125 - 100 - 75 - 50 - 25 - 0			125 - 100 - 75 - 50 - 25 - 0	o O and 1 are similar	1
:	<pre>ax.set_xla ax.set_yla</pre>	<pre>t.subplots(n s = zip(axs. in ax_title ot(y = n_var abel(f"{n_va} abel(None) 'Box Plots f remove()</pre>	rows=2, flat, nu _pairs: _,data = r}", fon or numer	ncols=4, fi m_var) df, ax = ax tsize = 15)	gsize = (15 x, color = '	,10))			
90 80 70 60 50			8000 - 7000 - 6000 - 5000 - 4000 - 2000 - 1000 -	.J(S	un	nerical va 80	6	00000 -	•
	ag	e		eatinine_phos	phokinase	ejection 250 - 200 - 150 - 100 -	n_fraction	pla	atelets
9]: [,	serum_credf[(df['creat: age anaemi 1 55.0 4 65.0 9 80.0	inine_phosph a creatinine_ph 0 1	okinase' nosphokina 78 1	diabetes 61 0 60 1 23 0	<pre>cdf['serum_c ejection_fraction 3 2 3</pre>	tireatinine']>8) n high_blood_pres	0 263358.03 0 327000.00 1 388000.00	erum_creatinine s 1.1 2.7 9.4	erum_sodiu 1: 1 1:
1 1 2 2 Oi	60 45.0 09 45.0 99 60.0 17 54.0 96 45.0 df = df[~((df	0 0 1 0 ['creatinine] n see the outlie	77 2 12 4 20 _phospho	02 1 92 1 11 1 27 0 60 1 kinase']>70 ye have outlier	2 3 3 7 6	5 5 0 0 serum_creatini	1 388000.00 1 390000.00 0 850000.00 0 263358.03 1 151000.00 0 742000.00 ne']>8) (df[1.0 1.3 1.8 9.0 0.8	1; 1, 1; 1; 00000) [
so 1]:	<pre>fig, axs = pla ax_title_pairs for ax, n_var sns.boxplo ax.set_xla ax.set_yla</pre>	o make it a mu t.subplots(n s = zip(axs. in ax_title ot(y = n_var abel(f"{n_va} abel(None) 'Box Plots f remove()	ch smaller rows=2, flat, nu _pairs: ,data = r}", fon or numer	ncols=4, fi m_var) df, ax = ax tsize = 15)	.gsize = (15 x, color = '	<pre>,10)) lawngreen') ize = 25, weig</pre>	ht = 'bold')		
90 80 70 60			6000 - 5000 - 4000 - 3000 - 2000 -	, Plots	r nun	nerical va 80 - 70 - 60 - 50 - 40 - 30 -	\$ 6 4 4 3 3 2 2	00000 -	
:	ag	e	145 - 140 - 135 - 125 -	eatinine_phos	phokinase	ejection 250 - 200 - 150 - 100 -	n_fraction	υ -L pla	atelets
Ak ou sig	df_grouped = df_grouped_d = fig, axs = plittitles2 = [x :	df.groupby(bedf.groupbyt.subplots(n	ting outlien dots has y=['sex'(by=['DErows=7, zip(num_	changed and]) ATH_EVENT'] ncols=2, fi var, num_var	ant to look on now these value) .gsize=(12,	Y axis, because the less above or below	ime nese dots on plots w horizontal lines	_	
	sns.d: ax.sei ax.leq else: sns.d: sns.d: ax.sei ax.leq i+=1	<pre>in ax_title : istplot(df_g istplot(df_g t_xlabel(f"{ gend(title=' istplot(df_g istplot(df_g t_xlabel(f"{ gend(title='</pre>	_pairs: rouped.g rouped.g title}", Sex') rouped_d rouped_d title}",	<pre>ret_group(0) ret_group(1) fontsize = l.get_group(l.get_group(fontsize =</pre>	<pre>[title], bi 15) (0) [title], ; (1) [title], ;</pre>	ns=10, ax=ax, bins=10, ax=ax	<pre>label='Woman', label='Man', c , label='No', , label='Yes',</pre>	<pre>color='red')</pre>	blue')
	0.035 - 0.030 - 0.025 - 20.020 - 0.015 - 0.010 - 0.005 -		50	80	Sex Woman Man	0.040 - 0.035 - 0.030 - 0.025 - 0.020 - 0.015 - 0.010 - 0.005 -	40 60	80	Death_Event No Yes
Density	0.0016 - 0.0014 - 0.0012 - 0.0008 - 0.0006 - 0.0004 - 0.0002 -		age		Sex Woman Man	0.0014 - 0.0012 - 0.0010 - 0.0008 - 0.0006 - 0.0004 - 0.0002 -		ige	Death_Event No Yes
	0.0000 -1000 0 0.007 - 0.006 - 0.005 - 2.004 - 0.003 - 0.002 - 0.001 -	creatinin		4000 5000 hokinase	Sex Woman Man	0.0000 -1000 0		hosphokinase	6000 700 Death_Event No Yes
	0.00 le-6 7- 6- 5- Aiya 4- 2- 1-		tion_frac	ction	Sex Woman Man	0.00 le-6 5 - 5 - 1 -	ejection	n_fraction	80 Death_Event No Yes
	1 - 0 - 1000000 0 0 0 0 0 0 0 0 0 0 0 0		300000 4000 platelets	000 500000 60000	Sex Woman Man	1.0 - -1000000 0		00 400000 500000 600 celets	Death_Event No Yes
	0.14 - 0.12 - 0.10 - 10 - 10 - 10 - 10 - 10 - 10 -	seru	m_creati	inine	Sex Woman Man	0.12 - 0.10 - 0.08 - 0.06 - 0.04 -	serum_c	creatinine	Death_Event No Yes
	0.007 - 0.006 - 0.005 - 0.004 - 0.003 - 0.002 -	20 125 13 ser	0 135 um_sodi	140 145 um	Sex Woman Man	0.002 - 0.0012 - 0.008 - 0.006 - 0.004 - 0.002		135 140 14 sodium	5 150 Death_Event No Yes
ev pe th	ent. There are or cople in the prob	distributions of ally little differer e had around 6 who died had h .subplots(fi	numeric vonces on the 50-70 year igher seru	e plots betweens. More peoplem creatinine.	he left we see o en woman and e who died we	man. From plots re older, had less	n gender, on the ron the right we can ejection fraction at 100 days of observ	n conclude more. nd serum sodium	The most
1 1 1	sns.heatmap(co	orrMatrix, a ', labelsize ', labelsize tation=45, h age -	=15) =15)				-0.8	3	
	ejection_frac high_blood_pres plate serum_creati serum_soc	ssure - elets - nine -					- 0.6 - 0.4 - 0.3 - 0.0	2	
4]:		time -	inase habetes had been had	n phood pressure date	ets string sodium	grating time Death	0		
4]:	ar reatinine_phospho di ejection_fi high_blood_pr	age 1.00000 naemia 0.07130 okinase -0.03554 abetes -0.08726 raction 0.07349	0.0713 0.0713 0.0000 0.0713 0.0000 0.01776 0.0005 0.0005 0.00266	808 900 870 881 825	_phosphokinase -0.035542 -0.177670 1.000000 -0.032887 -0.035565 -0.102230 -0.038080	-0.087263 0.000581 -0.032887 1.000000 0.007448	0.073494 0.029825 -0.035565 0.007448 1.000000 0.018585 0.077183	0.090114 -0.4 0.026674 -0.4 -0.102230 -0.4 -0.000905 0.4 0.018585 0.4 1.000000 0.4	008403 012700 038080 047503 077183 082537
sn	serum_crea serum_s sn DEATH_ is always recomn noking and sex, v	sex 0.06411 noking 0.122094 EVENT 0.25872	21 -0.0172 50 0.0588 19 -0.0848 95 -0.1080 40 -0.1388 26 0.0538 the correlation	286 314 316 396 374 325 ation matrix. N	0.028416 0.076827 0.061756 0.050642 0.048995 -0.001288	-0.010344 -0.067525 -0.160036 -0.154950 0.030021 0.007141 sions are low, below	0.077183 -0.082086 0.166332 -0.135108 -0.061515 0.010066 -0.276931 ow 0,2. The highest stinine with higher	-0.091433 -0.4 0.018756 0.4 -0.109399 -00.055570 0.4 -0.199059 0.4 0.063744 -0.4	055369 056887 144856 006640 009134 040354 ween
va [5]:	riables. from sklearn : from scipy imp df_mutual=df.d minfos=[] for var in car print("\n' print("Pea print("Mut minfos.app print("Ch:	<pre>import featu port stats copy() t_var[:-1]: ", var) arson", stat tual info", pend(feature i2", feature</pre>	re_selecti_selecti	onr(df_mutua selection.m on.mutual_i	al["DEATH_EV nutual_info_ .nfo_classif _mutual[var]	ENT"], df_mutu classif(df_mut (df_mutual[var .values.reshap		s.reshape(-1,1 pe(-1,1),df_mu ual["DEATH_EVE),df_mut tual["DE, NT"].val
P M C A	print("Ch: print("And imp = list(zi; imp.sort(reve: imp anaemia earson (0.053 utual info [0 hi2 (array([0 nova (array([diabetes earson (0.007 utual info [0	i2", feature ova", featur o(minfos, ca rse=True) 824963379838 .00715718] .4780259]), 0.83970234]) 140928985947	selecti e_select t_var)) 615, 0.3 array([0, array(360245929390 0.48931786]) ([0.36024593	_mutual[var] sif(df_mutua) 00044) 0) 0) 0) 0) 0)	.values.reshap	e(-1,1),df_mut	ual["DEATH_EVE	NT"].valı
M C A P M C A	utual info [0 hi2 (array([0 nova (array([high_blood_pr earson (0.063 utual info [0 hi2 (array([0 nova (array([sex earson (-0.00 utual info [0 hi2 (array([0 nova (array([sex earson (-sex earson (-se] .00871978]), 0.01473769]) essure 743755240603 .02968227] .76795734]), 1.17907487]) 160160189046] .00026164]), 0.00074132])	array([, array() array() array() array() array() array() array()	[0.92560187] ([0.90345946 449309598181 [0.38084971] ([0.27844931 0.9782972591 ([0.97829726	[])) 5])) 176) [])) 13)) 1621898)				
P M M C A A 55]: [earson (-0.00 utual info [0 hi2 (array([0 nova (array([(array([0.041 (array([0.036 (array([0]), (array([0]), minfos=[] for var in nur print(var minfos.app] .0003403]), 0.00049671]) 5939]), 'sex 09193]), 'sm 4272]), 'hig 'diabetes'), 'anaemia')] m_var: , " Mutual i pend(feature	array([(, array(i'), doking'), th_blood_ nfo", fe _selecti	0.98528198]) ([0.98223443	ction.mutual		df[var].astype		
a c e p s s t t	<pre>imp = list(zigoimp.sort(reversimp) ge Mutual in reatinine_pho jection_fract latelets Mut erum_creatini erum_sodium ime Mutual i (array([0.244 (array([0.054 (array([0.054 (array([0.045</pre>	fo [0.066632 sphokinase ion Mutual ual info [0] ne Mutual i Mutual info nfo [0.23828 76638]), 'ti 72493]), 'ej 30751]), 'se	m_var)) 41] Mutual i info [0.0 [0.02381 43] me'), ection_ferum_crea	.nfo [0.0171 .07141664] .06271535] .683]	_ [6831]	~yp	_ ue	-,1	نگ ر
In we ca be	(array([0.045 (array([0.006 (array([0]), order to choose eren't problems to tegorical variable ecause they are signature.	90757]), 'ag 4123]), 'cre 72656]), 'se 'platelets') properly variab pecause in mos es it wasn't so e ignificant in cas el we use data	re'), eatinine_erum_sodi bles, we ap t of cases, easy because of inform	phosphokina and the phosphokina and the previous in the previous and the p	selection meth nation for plate nutual informa respondents h s part and ada	elets was 0 and th tion had changed ealth. pt it to be used in	ode many times. Wis variable is not used to so we decided to the model. For the which approach is to	sed in modelling. use all the catego	With orical varial ndardize d
7]:	<pre>from sklearn.n from sklearn.n f</pre>	model_select metrics impo preprocessin linear_model svm import S tree import ensemble imp import XGBC1 moking','hig ime','ejecti	ion impo rt confu g import import VC Decision ort Rand assifier h_blood_ on_fract	rt train_te sion_matrix MinMaxScal LogisticReg TreeClassif domForestCla pressure',' ion','serum	est_split,cr x, accuracy_ er gression fier assifier sex', 'diab a_creatinine	oss_validate, score, precisi	cross_val_scor on_score, reca	e#, GridSearch ll_score, f1_s	CV Core, ro
3]:	X = df.drop([' y = df["DEATH]	"DEATH_EVENT _EVENT"] axScaler() scaler.fit_t	", 'plat	elets'], ax	ris=1)		.3, random_sta		
3]:	_	st, y_train,	y_test						

