Nama: Rafa Al Razzak

NIM: 0110224155

1. Analisis data tersebut (ukuran dataset, type data dari tiap variabel, apakah terdapat missing values, data yang duplikat, dan outlier) dan berikan kesimpulan terkait apakah data tersebut dapat langsung diolah!

Jawab

- Dataset berisi seribu tiga ratus empat puluh baris dan tujuh kolom. Variabel umur dan jumlah anak sudah berupa angka, sedangkan jenis kelamin, status merokok, dan wilayah berupa teks.
- Dua kolom penting yaitu indeks massa tubuh dan biaya asuransi seharusnya numerik, tetapi pada data mentah masih tersimpan sebagai teks sehingga perlu dikonversi terlebih dahulu.
- Terdapat sedikit nilai hilang pada indeks massa tubuh dan biaya asuransi, serta ada tiga baris data yang terduplikasi.
- Pemeriksaan dengan boxplot menunjukkan adanya pencilan terutama pada biaya asuransi, yang wajar pada data biaya namun tetap perlu diperhatikan saat analisis.
- Dengan kondisi tersebut, data belum dapat langsung diolah. Data perlu dibersihkan dulu dengan menghapus duplikasi, memperbaiki tipe data indeks massa tubuh dan biaya asuransi menjadi numerik, menangani nilai hilang dengan cara yang sesuai, dan meninjau dampak pencilan sebelum melanjutkan ke tahap analisis berikutnya.

2. Buatlah data yang bersih dengan menghapus data-data yang bermasalah dan lampirkan hasil data yang sudah dibersihkan!

age	sex	bmi	children	smoker	region	charges
18	male	33,77	1	no	southeast	17255523
28	male	33	3	no	southeast	4449462
33	male	22705	0	no	northwest	2,2E+09
32	male	28,88	0	no	northwest	38668552
31	female	25,74	0	no	southeast	37566216
46	female	33,44	1	no	southeast	82405896
37	female	27,74	3	no	northwest	72815056
37	male	29,83	2	no	northeast	64064107
60	female	25,84	0	no	northwest	2,89E+09
25	male	26,22	0	no	northeast	27213208

62	female	26,29	0	yes	southeast	2,78E+08
23	male	34,4	0	no	southwest	1826843
56	female	39,82	0	no	southeast	1,11E+08
27	male	42,13	0	yes	southeast	3,96E+08
52	female	30,78	1	no	northeast	1,08E+08
23	male	23845	0	no	northeast	2,4E+08
56	male	40,3	0	no	southwest	10602385
30	male	35,3	0	yes	southwest	36837467
60	female	36005	0	no	northeast	1,32E+09
30	female	32,4	1	no	southwest	4149736
18	male	34,1	0	no	southeast	1137011
34	female	31,92	1	yes	northeast	3,77E+08
37	male	28025	2	no	northwest	6,2E+08
59	female	27,72	3	no	southeast	1,4E+08
63	female	23085	0	no	northeast	1,45E+09
55	female	32775	2	no	northwest	1,23E+09
23	male	17385	1	no	northwest	2,78E+08
31	male	36,3	2	yes	southwest	38711
22	male	35,6	0	yes	southwest	35585576
18	female	26315	0	no	northeast	2,2E+08
63	male	28,31	0	no	northwest	1,38E+08
28	male	36,4	1	yes	southwest	5,12E+09
19	male	20425	0	no	northwest	1,63E+08
62	female	32965	3	no	northwest	1,56E+09
35	male	36,67	1	yes	northeast	3,98E+08
60	male	39,9	0	yes	southwest	48173361
31	female	36,63	2	no	southeast	49497587
41	male	21,78	1	no	southeast	62724772
38	male	37,05	1	no	northeast	60796715
55	male	37,3	0	no	southwest	2,06E+09
18	female	38665	2	no	northeast	3,39E+08
28	female	34,77	0	no	northwest	35569223
60	female	24,53	0	no	southeast	1,26E+08
36	male	35,2	1	yes	southeast	38709176
18	female	35625	0	no	northeast	2,21E+08
21	female	33,63	2	no	northwest	35798287
48	male	28	1	yes	southwest	23568272
36	male	34,43	0	yes	southeast	3,77E+08
40	female	28,69	3	no	northwest	80596791
58	male	36955	2	yes	northwest	4,75E+09
58	female	31825	2	no	northeast	1,36E+09
18	male	31,68	2	yes	southeast	3,43E+08
53	female	22,88	1	yes	southeast	2,32E+08
34	female	37335	2	no	northwest	5,99E+08
43	male	27,36	3	no	northeast	86062174
25	male	33,66	4	no	southeast	45046624
28	female	25935	1	no	northwest	4,13E+08
20	female	22,42	0	yes	northwest	1,47E+08

61	female	39,1	2	no	southwest	14235072
40	male	26315	1	no	northwest	6,39E+08
40	female	36,19	0	no	southeast	59201041
28	male	23,98	3	yes	southeast	1,77E+08
27	female	24,75	0	yes	southeast	1,66E+08
58	male	32,01	1	no	southeast	1,19E+08
57	male	34,01	0	no	northwest	1,14E+08
29	female	29,59	1	no	southeast	39474131
21	male	35,53	0	no	southeast	15324697
22	female	39805	0	no	northeast	2,76E+08
41	female	32965	0	no	northwest	6,57E+08
31	male	26885	1	no	northeast	4,44E+08
45	female	38285	0	no	northeast	7,94E+08
22	male	37,62	1	yes	southeast	3,72E+08
48	female	41,23	4	no	northwest	1,1E+08
37	female	34,8	2	yes	southwest	39836519
45	male	22895	2	yes	northwest	2,11E+09
57	female	31,16	0	yes	northwest	4,36E+08
46	female	27,74	0	no	northwest	80266666
55	female	26,98	0	no	northwest	1,11E+08
21	female	39,49	0	no	southeast	20269741
53	female	24795	1	no	northwest	1,09E+09
59	male	29,83	3	yes	northeast	3,02E+08
35	male	34,77	2	no	northwest	57290053
28	female	37,62	1	no	southeast	37668838
55	male	38,28	0	no	southeast	1,02E+08
56	male	19,95	0	yes	northeast	2,24E+08
41	female	31,6	0	no	southwest	6186127
30	male	25,46	0	no	northeast	36450894
18	female	30115	0	no	northeast	2,13E+08
61	female	29,92	3	yes	southeast	3,09E+08
20	male	28025	1	yes	northwest	1,76E+09
26	male	30875	2	no	northwest	3,88E+08
29	male	27,94	0	no	southeast	28671196
63	male	35,09	0	yes	southeast	4,71E+08
54	male	33,63	1	no	northwest	1,08E+08
21	female	35,72	0	no	northwest	24047338
52	male	32205	3	no	northeast	1,15E+09
60	male	28595	0	no	northeast	3,03E+09
58	male	49,06	0	no	southeast	1,14E+08
29	female	27,94	1	yes	southeast	1,91E+08
49	female	27,17	0	no	southeast	86013293
37	female	23,37	2	no	northwest	66864313
44	male	37,1	2	no	southwest	7740337
18	male	23,75	0	no	northeast	17056245
20	female	28975	0	no	northwest	2,26E+08
44	male	31,35	1	yes	northeast	3,96E+08
47	female	33915	3	no	northwest	1,01E+09

26	female	28785	0	no	northeast	3,39E+08
52	female	37,4	0	no	southwest	9634538
32	female	17765	2	yes	northwest	3,27E+08
38	male	34,7	2	no	southwest	6082405
59	female	26505	0	no	northeast	1,28E+09
53	female	35,9	2	no	southwest	11163568
19	male	25555	0	no	northwest	1,63E+08
20	female	28785	0	no	northeast	2,46E+08
19	male	34,1	0	no	southwest	1261442
22	male	25175	0	no	northwest	2,05E+08
54	female	31,9	3	no	southeast	2,73E+09
22	female	36	0	no	southwest	2166732
34	male	22,42	2	no	northeast	2,74E+09
26	male	32,49	1	no	northeast	34905491
29	male	29735	2	no	northwest	18157876
30	male	28,69	3	yes	northwest	2,07E+08
29	female	38,83	3	no	southeast	51382567
46	male	30495	3	yes	northwest	4,07E+09
51	female	37,73	1	no	southeast	98776077
53	female	37,43	1	no	northwest	1,1E+08
35	male	24,13	1	no	northwest	51252157
32	female	37145	3	no	northeast	6,33E+08
42	female	23,37	0	yes	northeast	2E+08
40	female	25,46	1	no	northeast	70771894
44	male	39,52	0	no	northwest	69487008
48	male	24,42	0	yes	southeast	2,12E+08
18	male	25175	0	yes	northeast	1,55E+09
30	male	35,53	0	yes	southeast	3,7E+08
50	female	27,83	3	no	southeast	1,97E+09
18	female	36,85	0	yes	southeast	3,61E+08
54	male	39,6	1	no	southwest	10450552
37	male	29,64	0	no	northwest	50281466
47	male	28215	4	no	northeast	1,04E+09
20	female	37	5	no	southwest	4830,63
32	female	33155	3	no	northwest	6,13E+08
19	female	31825	1	no	northwest	2,72E+08
27	male	18905	3	no	northeast	4,83E+08
63	male	41,47	0	no	southeast	1,34E+08
18	male	15,96	0	no	northeast	16947964
35	female	34,8	1	no	southwest	5246047
24	female	33345	0	no	northwest	2,86E+08
63	female	37,7	0	yes	southwest	48824,45
38	male	27835	2	no	northwest	6,46E+08
54	male	29,2	1	no	southwest	10436096
41	female	33155	3	no	northeast	8,54E+08
58	male	28595	0	no	northwest	1,17E+09
18	female	38,28	0	no	southeast	16318212
22	male	19,95	3	no	northeast	40054225

44	female	26,41	0	no	northwest	74194779
44	male	30,69	2	no	southeast	77314271
36	male	41895	3	yes	northeast	4,38E+09
26	female	29,92	2	no	southeast	39819768
41	female	32,2	1	no	southwest	6775961
29	female	32,11	2	no	northwest	49229159
61	male	31,57	0	no	southeast	1,26E+08
25	male	25,74	0	no	southeast	21376536
18	male	34,43	0	no	southeast	11374697
19	male	30,59	0	no	northwest	16395631
39	female	32,8	0	no	southwest	5649715
64	female	39,33	0	no	northeast	1,49E+08
19	female	32,11	0	no	northwest	21306759
48	female	32,23	1	no	southeast	88711517
60	female	24035	0	no	northwest	1,3E+09
27	female	36,08	0	yes	southeast	3,71E+08
28	female	28,88	1	no	northeast	43377352
35	male	27,74	2	yes	northeast	2,1E+08
40	male	41,23	1	no	northeast	66101097
40	male	30875	4	no	northwest	8,16E+08
34	female	26,73	1	no	southeast	50027827
41	female	37,1	2	no	southwest	7371772
26	female	29,92	1	no	southeast	33929768
24	female	23,21	0	no	southeast	2,51E+09
34	female	33,7	1	no	southwest	5012471
53	female	33,25	0	no	northeast	1,06E+08
19	male	34,8	0	yes	southwest	34779615
42	male	24,64	0	yes	southeast	1,95E+08
55	male	33,88	3	no	southeast	1,2E+08
28	male	38,06	0	no	southeast	26894954
58	female	41,91	0	no	southeast	2,42E+09
41	female	31635	1	no	northeast	7,36E+08
47	male	25,46	2	no	northeast	92252564
42	female	36195	1	no	northwest	7,44E+08
59	female	27,83	3	no	southeast	1,4E+08
39	male	24,51	2	no	northwest	67101919
40	female	22,22	2	yes	southeast	1,94E+08
18	female	26,73	0	no	southeast	16157667
31	male	38,39	2	no	southeast	44632051
44	male	38,06	1	no	southeast	71526714
23	female	36,67	2	yes	northeast	3,85E+08
33	female	22135	1	no	northeast	5,35E+08
40	male	35,3	3	no	southwest	7196867
63	female	27,74	0	yes	northeast	2,95E+08
54	male	30,02	0	no	northwest	2,45E+09
60	female	38,06	0	no	southeast	1,26E+08
24	male	35,86	0	no	southeast	19869334
29	male	28975	1	no	northeast	4,04E+08

18	male	17,29	2	yes	northeast	1,28E+08
63	female	32,2	2	yes	southwest	47305305
54	male	34,21	2	yes	southeast	4,43E+08
50	male	31825	0	yes	northeast	4,11E+09
55	female	25365	3	no	northeast	1,3E+09
56	male	33,63	0	yes	northwest	4,39E+08
38	female	40,15	0	no	southeast	54009805
51	male	24415	4	no	northwest	1,15E+09
19	male	31,92	0	yes	northwest	3,38E+08
20	female	26,84	1	yes	southeast	1,71E+08
52	male	24,32	3	yes	northeast	2,49E+08
19	male	36955	0	yes	northwest	3,62E+09
53	female	38,06	3	no	southeast	2,05E+09
46	male	42,35	3	yes	southeast	4,62E+08
59	female	32395	3	no	northeast	1,46E+09
45	male	30,2	1	no	southwest	7441053
49	male	25,84	1	no	northeast	92824806
18	male	29,37	1	no	southeast	17194363
50	male	34,2	2	yes	southwest	42856838
41	male	37,05	2	no	northwest	72657025
50	male	27455	1	no	northeast	9,62E+08
25	male	27,55	0	no	northwest	25231695
19	male	20615	2	no	northwest	2,8E+08
59	male	31,79	2	no	southeast	1,29E+08
51	female	21,56	1	no	southeast	98551314
54	male	40565	3	yes	northeast	4,85E+09
30	male	27645	1	no	northeast	4,24E+08
55	female	32395	1	no	northeast	1,19E+09
52	female	31,2	0	no	southwest	9625,92
46	male	26,62	1	no	southeast	77421098
46	female	48,07	2	no	northeast	94329253
63	female	26,22	0	no	northwest	1,43E+08
59	female	36765	1	yes	northeast	4,79E+09
28	female	33,4	0	no	southwest	3172018
29	male	29,64	1	no	northeast	2,03E+09
25	male	45,54	2	yes	southeast	4,21E+08
22	female	28,82	0	no	southeast	21567518
18	male	22,99	0	no	northeast	17045681
47	male	25,41	1	yes	southeast	2,2E+08
31	male	34,39	3	yes	northwest	3,87E+08
48	female	28,88	1	no	northwest	92494952
36	male	27,55	3	no	northeast	67467425
53	female	22,61	3	yes	northeast	2,49E+08
56	female	37,51	2	no	southeast	1,23E+08
28	female	33	2	no	southeast	4349462
57	female	38	2	no	southwest	12646207
29	male	33345	2	no	northwest	1,94E+08
30	female	33,33	1	no	southeast	41510287

58	male	34865	0	no	northeast	1,19E+09
41	female	33,06	2	no	northwest	77491564
43	male	35,97	3	yes	southeast	4,21E+08
49	male	35,86	0	no	southeast	81244084
27	female	31,4	0	yes	southwest	34838873
52	male	33,25	0	no	northeast	97227695
50	male	32205	0	no	northwest	8,84E+08
54	male	32775	0	no	northeast	1,04E+09
44	female	27645	0	no	northwest	7,42E+08
32	male	37335	1	no	northeast	4,67E+08
34	male	25,27	1	no	northwest	48947533
26	female	29,64	4	no	northeast	2,47E+09
57	male	40945	0	no	northeast	1,16E+09
40	male	34105	1	no	northeast	6,6E+08
27	female	23,21	1	no	southeast	35618889
45	male	36,48	2	yes	northwest	4,28E+08
64	female	33,8	1	yes	southwest	47928,03
52	male	36,7	0	no	southwest	9144565
61	female	36385	1	yes	northeast	4,85E+09
52	male	27,36	0	yes	northwest	2,44E+08
61	female	31,16	0	no	northwest	1,34E+08
56	female	28785	0	no	northeast	1,17E+09
43	female	35,72	2	no	northeast	1,91E+09
64	male	34,5	0	no	southwest	13822803
60	male	25,74	0	no	southeast	1,21E+08
62	male	27,55	1	no	northwest	1,39E+08
50	male	32,3	1	yes	northeast	41919097
46	female	27,72	1	no	southeast	82326388
62	male	30,02	0	no	northwest	1,34E+08
60	female	27,55	0	no	northeast	1,32E+08
63	male	36765	0	no	northeast	1,4E+09
49	female	41,47	4	no	southeast	1,1E+08
34	female	29,26	3	no	southeast	61842994
33	male	35,75	2	no	southeast	48899995
46	male	33345	1	no	northeast	8,33E+08
36	female	29,92	1	no	southeast	54780368
19	male	27835	0	no	northwest	1,64E+08
57	female	23,18	0	no	northwest	1,18E+08
33	male	35245	0	no	northeast	1,24E+08
18	female	38,28	0	no	southeast	1,41E+09
46	male	43,89	3	no	southeast	89441151
47	male	29,83	3	no	northwest	96203307
23	male	41,91	0	no	southeast	18372819
18	female	20,79	0	no	southeast	16075101
48	female	32,3	2	no	northeast	10043249
21	female	21,89	2	no	southeast	31805101
49	female	30,78	1	no	northeast	97783472
56	female	32,3	3	no	northeast	13430265

42	female	24985	2	no	northwest	8,02E+08
57	female	22,23	0	no	northeast	1,2E+08
42	female	33155	1	no	northeast	7,64E+08
26	male	32,9	2	yes	southwest	36085219
20	male	33,33	0	no	southeast	13915287
23	female	28,31	0	yes	northwest	1,8E+08
39	female	24,89	3	yes	northeast	2,17E+08
24	male	40,15	0	yes	southeast	3,81E+08
64	female	30115	3	no	northwest	1,65E+09
62	male	31,46	1	no	southeast	2,7E+09
27	female	17955	2	yes	northeast	1,5E+09
55	male	30685	0	yes	northeast	4,23E+09
55	male	33	0	no	southeast	2,08E+09
35	female	43,34	2	no	southeast	58469176
44	male	22135	2	no	northeast	8,3E+08
19	male	34,4	0	no	southwest	1261859
58	female	39,05	0	no	southeast	1,19E+08
50	male	25365	2	no	northwest	3,03E+09
26	female	22,61	0	no	northwest	31768159
24	female	30,21	3	no	northwest	46180799
48	male	35625	4	no	northeast	1,07E+09
19	female	37,43	0	no	northwest	21380707
48	male	31445	1	no	northeast	8,96E+08
49	male	31,35	1	no	northeast	92901395
46	female	32,3	2	no	northeast	9411005
46	male	19855	0	no	northwest	7,53E+08
43	female	34,4	3	no	southwest	8522003
21	male	31,02	0	no	southeast	1,66E+09
18	female	38,17	0	no	southeast	16316683
47	male	47,52	1	no	southeast	80839198
64	female	32965	0	no	northwest	1,47E+09
49	male	32,3	3	no	northwest	10269,46
52	female	38,38	2	no	northeast	1,14E+08
33	female	24,31	0	no	southeast	41850979
19	male	17,48	0	no	northwest	16213402
44	female	20235	1	yes	northeast	1,96E+09
26	female	17195	2	yes	northeast	1,45E+09
19	female	35,15	0	no	northwest	21349015
43	female	35,64	1	no	southeast	73457266
52	male	34,1	0	no	southeast	9140951
64	male	39,16	1	no	southeast	1,44E+08
63	female	26,98	0	yes	northwest	2,9E+08
64	male	33,88	0	yes	southeast	4,69E+08
61	male	35,86	0	yes	southeast	4,66E+08
40	male	32775	1	yes	northeast	3,91E+09
25	male	30,59	0	no	northeast	27273951
48	male	30,2	2	no	southwest	8968,33
45	male	24,31	5	no	southeast	97888659

38	female	27265	1	no	northeast	6,56E+08
18	female	29165	0	no	northeast	7,32E+09
21	female	16815	1	no	northeast	3,17E+08
19	male	33,1	0	no	southwest	2,31E+09
29	female	20235	2	no	northwest	4,91E+08
31	male	28595	1	no	northwest	4,24E+08
60	male	33,11	3	no	southeast	1,39E+08
22	male	31,73	0	no	northeast	22547967
52	female	46,75	5	no	southeast	1,26E+08
26	male	29,45	0	no	northeast	28973235
31	female	32,68	1	no	northwest	47382682
33	female	33,5	0	yes	southwest	37079372
18	male	43,01	0	no	southeast	11493959
59	female	36,52	1	no	southeast	2,83E+09
56	male	26695	1	yes	northwest	2,61E+09
45	female	33,1	0	no	southwest	7345084
60	male	29,64	0	no	northeast	1,27E+08
56	female	25,65	0	no	northwest	1,15E+08
35	male	38,6	1	no	southwest	4762329
30	male	24,13	1	no	northwest	40322407
20	male	29735	0	no	northwest	1,77E+08
32	male	46,53	2	no	southeast	46863887
59	male	37,4	0	no	southwest	2,18E+08
55	female	30,14	2	no	southeast	1,19E+08
57	female	30495	0	no	northwest	1,18E+09
56	male	39,6	0	no	southwest	10601412
40	female	33	3	no	southeast	7682,67
49	female	36,63	3	no	southeast	1,04E+08
42	male	30	0	yes	southwest	22144032
62	female	38095	2	no	northeast	1,52E+09
56	male	25935	0	no	northeast	1,12E+09
19	male	25175	0	no	northwest	1,63E+08
30	female	28,38	1	yes	southeast	1,95E+08
56	female	33,82	2	no	northwest	1,26E+08
28	female	24,32	1	no	northeast	2,33E+08
27	male	32,67	0	no	southeast	24970383
18	female	30115	0	no	northeast	2,2E+08
47	female	33345	0	no	northeast	2,09E+09
61	male	28,31	1	yes	northwest	2,89E+08
25	male	35625	0	no	northwest	2,53E+08
21	male	36,85	0	no	southeast	15343045
23	male	32,56	0	no	southeast	18242854
63	male	41325	3	no	northwest	1,56E+09
49	male	37,51	2	no	southeast	93047019
18	female	31,35	0	no	southeast	16221885
51	female	39,5	1	no	southwest	9880068
48	male	34,3	3	no	southwest	9563029
31	female	31065	0	no	northeast	4,35E+08

54	female	21,47	3	no	northwest	1,25E+08
44	female	38,06	0	yes	southeast	4,89E+09
53	male	31,16	1	no	northwest	1,05E+08
19	female	32,9	0	no	southwest	1748774
61	male	43,4	0	no	southwest	12574049
20	male	27,93	0	no	northeast	19670227
44	female	23,98	2	no	southeast	82111002
62	female	39,2	0	no	southwest	13470,86
29	male	34,4	0	yes	southwest	36197699
51	male	23,21	1	yes	southeast	2,22E+08
19	male	30,25	0	yes	southeast	3,25E+08
38	female	28,93	1	no	southeast	59743847
37	male	30875	3	no	northwest	6,8E+08
22	male	31,35	1	no	northwest	26432685
21	male	23,75	2	no	northwest	30770955
24	female	25,27	0	no	northeast	30442133
56	male	32,11	1	no	northeast	1,18E+08
27	male	33,66	0	no	southeast	24984144
51	male	22,42	0	no	northeast	93613268
58	male	35,7	0	no	southwest	11362755
20	male	35,31	1	no	southeast	2,77E+09
45	male	30495	2	no	northwest	8,41E+08
35	female	31	1	no	southwest	5240765
31	male	30875	0	no	northeast	3,86E+08
50	female	27,36	0	no	northeast	2,57E+09
32	female	44,22	0	no	southeast	39941778
51	female	33915	0	no	northeast	9,87E+08
38	female	37,73	0	no	southeast	53976167
18	female	33,88	0	no	southeast	1,15E+09
19	female	30,59	2	no	northwest	2,41E+09
46	male	39425	1	no	northeast	8,34E+08
18	male	25,46	0	no	northeast	17080014
57	male	42,13	1	yes	southeast	4,87E+08
62	female	31,73	0	no	northeast	1,4E+08
37	male	36,19	0	no	southeast	1,92E+09
64	male	40,48	0	no	southeast	1,38E+08
38	male	28025	1	no	northeast	6,07E+08
33	female	38,9	3	no	southwest	5972378
46	female	30,2	2	no	southwest	8825086
53	male	31,35	0	no	southeast	2,73E+09
34	female	38	3	no	southwest	6196448
20	female	31,79	2	no	southeast	30563881
63	female	36,3	0	no	southeast	13887204
54	female	47,41	0	yes	southeast	6,38E+09
54	male	30,21	0	no	northwest	1,02E+08
49	male	25,84	2	yes	northwest	2,38E+08
28	male	35435	0	no	northeast	3,27E+08
54	female	46,7	2	no	southwest	11538421

25	female	28595	0	no	northeast	3,21E+08
43	female	46,2	0	yes	southeast	45863205
32	female	28,93	0	no	southeast	39729247
52	female	31,73	2	no	northwest	1,12E+08
25	female	41325	0	no	northeast	1,79E+09
46	male	33,44	1	no	northeast	83345896
34	male	34,21	0	no	southeast	39351799
35	female	34105	3	yes	northwest	4E+09
19	male	35,53	0	no	northwest	16464297
46	female	19,95	2	no	northwest	91938385
54	female	32,68	0	no	northeast	1,09E+08
50	male	44,77	1	no	southeast	90587303
18	female	32,12	2	no	southeast	28012588
19	female	30495	0	no	northwest	2,13E+08
38	female	40565	1	no	northwest	6,37E+08
41	male	30,59	2	no	northwest	72567231
49	female	31,9	5	no	southwest	11552904
48	male	40565	2	yes	northwest	4,57E+09
18	female	37,29	1	no	southeast	22194451
30	female	43,12	2	no	southeast	47536368
62	female	36,86	1	no	northeast	3,16E+09
57	female	34295	2	no	northeast	1,32E+09
58	female	27,17	0	no	northwest	1,22E+08
22	male	26,84	0	no	southeast	16649996
31	female	38095	1	yes	northeast	5,86E+09
52	male	30,2	1	no	southwest	9724,53
25	female	23465	0	no	northeast	3,21E+08
59	male	25,46	1	no	northeast	1,29E+08
39	male	45,43	2	no	southeast	63562707
32	female	23,65	1	no	southeast	1,76E+09
33	female	28,27	1	no	southeast	47796023
21	male	20235	3	no	northeast	3,86E+08
34	female	30,21	1	yes	northwest	4,39E+08
61	female	35,91	0	no	northeast	1,36E+08
38	female	30,69	1	no	southeast	59768311
58	female	29	0	no	southwest	11842442
47	male	19,57	1	no	northwest	84280693
20	male	31,13	2	no	southeast	25664707
21	female	21,85	1	yes	northeast	1,54E+08
41	male	40,26	0	no	southeast	57091644
46	female	33725	1	no	northeast	8,82E+08
42	female	29,48	2	no	southeast	76403092
34	female	33,25	1	no	northeast	55948455
43	male	32,6	2	no	southwest	7441501
52	female	37525	2	no	northwest	3,35E+09
18	female	39,16	0	no	southeast	16330444
51	male	31635	0	no	northwest	9,17E+08
64	female	39,05	3	no	southeast	1,61E+08

19	female	28,31	0	yes	northwest	1,75E+08
51	female	34,1	0	no	southeast	9283562
27	female	25175	0	no	northeast	3,56E+08
59	female	23655	0	yes	northwest	2,57E+09
28	male	26,98	2	no	northeast	44350942
30	male	37,8	2	yes	southwest	39241442
47	female	29,37	1	no	southeast	85476913
38	female	34,8	2	no	southwest	6571544
18	female	33155	0	no	northeast	2,21E+08
34	female	19	3	no	northeast	6753038
20	female	33	0	no	southeast	1880,07
47	female	36,63	1	yes	southeast	4,3E+08
56	female	28595	0	no	northeast	1,17E+09
19	female	33,11	0	yes	southeast	3,44E+08
55	female	37,1	0	no	southwest	10713644
30	male	31,4	1	no	southwest	3659346
37	male	34,1	4	yes	southwest	40182246
18	male	33535	0	yes	northeast	3,46E+09
59	male	28785	0	no	northwest	1,21E+09
36	male	28,88	3	no	northeast	67485912
33	male	42,46	1	no	southeast	1,13E+09
58	male	38	0	no	southwest	11365952
44	female	38,95	0	yes	northwest	4,3E+08
53	male	36,1	1	no	southwest	10085846
29	female	35,53	0	no	southeast	33666697
40	male	22705	2	no	northeast	7,17E+08
51	male	39,7	1	no	southwest	9391346
64	male	38,19	0	no	northeast	1,44E+08
19	female	24,51	1	no	northwest	27091119
35	female	38095	2	no	northeast	2,49E+09
39	male	26,41	0	yes	northeast	2,01E+08
56	male	33,66	4	no	southeast	1,29E+08
33	male	42,4	5	no	southwest	6666243
42	male	28,31	3	yes	northwest	3,28E+09
61	male	33915	0	no	northeast	1,31E+09
23	female	34,96	3	no	northwest	44666214
43	male	35,31	2	no	southeast	1,88E+09
48	male	30,78	3	no	northeast	1,01E+08
39	male	26,22	1	no	northwest	61235688
40	female	23,37	3	no	northeast	82522843
58	female	32965	0	no	northeast	1,24E+09
49	female	42,68	2	no	southeast	98008882
53	female	39,6	1	no	southeast	10579711
48	female	31,13	0	no	southeast	82806227
45	female	36,3	2	no	southeast	8527532
59	female	35,2	0	no	southeast	12244531
26	female	42,4	1	no	southwest	3410324
27	male	33155	2	no	northwest	4,06E+08

48	female	35,91	1	no	northeast	2,64E+09
57	female	28785	4	no	northeast	1,44E+09
37	male	46,53	3	no	southeast	64356237
57	female	23,98	1	no	southeast	2,22E+09
32	female	31,54	1	no	northeast	51485526
18	male	33,66	0	no	southeast	11363994
64	female	22,99	0	yes	southeast	2,7E+08
43	male	38,06	2	yes	southeast	4,26E+08
40	female	32775	2	yes	northwest	4E+09
62	male	32015	0	yes	northeast	4,57E+09
40	female	29,81	1	no	southeast	65002359
30	male	31,57	3	no	southeast	48375823
29	female	31,16	0	no	northeast	39435954
41	female	31,02	0	no	southeast	61853208
44	female	43,89	2	yes	southeast	4,62E+08
45	male	21375	0	no	northwest	7,22E+08
55	female	40,81	3	no	southeast	1,25E+08
60	male	31,35	3	yes	northwest	4,61E+08
56	male	36,1	3	no	southwest	12363547
49	female	23,18	2	no	northwest	1,02E+08
39	male	35,3	2	yes	southwest	40103,89
53	male	24,32	0	no	northwest	98634718
53	male	26,41	2	no	northeast	1,12E+08
42	male	26125	2	no	northeast	7,73E+08
40	male	41,69	0	no	southeast	54387491
27	male	31,13	1	yes	southeast	3,48E+08
21	male	27,36	0	no	northeast	21041134
47	male	36,2	1	no	southwest	8068185
20	male	32395	1	no	northwest	2,36E+08
24	male	23655	0	no	northwest	2,35E+08
27	female	34,8	1	no	southwest	3577999
26	female	40185	0	no	northwest	3,2E+08
53	female	32,3	2	no	northeast	2,92E+09
41	male	35,75	1	yes	southeast	4,03E+08
56	male	33725	0	no	northwest	1,1E+09
23	female	39,27	2	no	southeast	35006123
21	female	34,87	0	no	southeast	20205523
50	female	44745	0	no	northeast	9,54E+08
53	male	41,47	0	no	southeast	95043103
34	female	26,41	1	no	northwest	53853379
47	female	29545	1	no	northwest	8,93E+08
33	female	32,9	2	no	southwest	5375038
51	female	38,06	0	yes	southeast	4,44E+08
49	male	28,69	3	no	northwest	1,03E+08
31	female	30495	3	no	northeast	6,11E+08
36	female	27,74	0	no	northeast	54690066
18	male	35,2	1	no	southeast	1727,54
50	female	23,54	2	no	southeast	1,01E+08

43	female	30685	2	no	northwest	8,31E+08
20	male	40,47	0	no	northeast	19844533
49	female	22,61	1	no	northwest	95669909
60	male	24,32	1	no	northwest	1,31E+08
51	female	36,67	2	no	northwest	1,08E+08
58	female	33,44	0	no	northwest	1,22E+08
51	female	40,66	0	no	northeast	98756804
53	male	36,6	3	no	southwest	11264541
62	male	37,4	0	no	southwest	12979358
19	male	35,4	0	no	southwest	1263249
50	female	27075	1	no	northeast	1,01E+09
30	female	39,05	3	yes	southeast	4,09E+08
41	male	28405	1	no	northwest	6,66E+08
29	female	21755	1	yes	northeast	1,67E+09
18	female	40,28	0	no	northeast	22176012
41	female	36,08	1	no	southeast	67813542
35	male	24,42	3	yes	southeast	1,94E+08
48	female	27265	1	no	northeast	9,45E+08
59	female	32,1	3	no	southwest	14007222
49	female	34,77	1	no	northwest	95838933
37	female	38,39	0	yes	southeast	4,04E+08
23	male	31,73	3	yes	northeast	3,62E+08
29	male	35,5	2	yes	southwest	4,46E+09
45	male	24035	2	no	northeast	8,6E+08
27	male	29,15	0	yes	southeast	1,82E+08
53	male	34105	0	yes	northeast	4,33E+09
31	female	26,62	0	no	southeast	37578448
50	male	26,41	0	no	northwest	88272099
50	female	30115	1	no	northwest	9,91E+08
34	male	27	2	no	southwest	1,17E+09
19	male	21755	0	no	northwest	1,63E+08
47	female	36	1	no	southwest	8556907
28	male	30875	0	no	northwest	3,06E+08
21	male	28975	0	no	northwest	1,91E+08
64	male	37905	0	no	northwest	1,42E+09
58	female	22,77	0	no	southeast	1,18E+08
24	male	33,63	4	no	northeast	1,71E+09
31	male	27645	2	no	northeast	5,03E+08
47	female	27,83	0	yes	southeast	2,31E+08
30	male	37,43	3	no	northeast	54287277
18	male	38,17	0	yes	southeast	3,63E+08
22	female	34,58	2	no	northeast	39257582
23	male	35,2	1	no	southwest	2416955
57	female	31825	0	no	northwest	1,18E+09
47	male	32,3	1	no	southwest	8062764
42	female	29	1	no	southwest	7050642
64	female	39,7	0	no	southwest	14319031
38	female	19475	2	no	northwest	6,93E+08

61	male	36,1	3	no	southwest	2,79E+09
44	female	36,48	0	no	northeast	1,28E+09
19	female	28,88	0	yes	northwest	1,77E+08
41	male	34,2	2	no	northwest	7261741
51	male	33,33	3	no	southeast	1,06E+08
40	male	32,3	2	no	northwest	6986697
45	male	39805	0	no	northeast	7,45E+08
35	male	34,32	3	no	southeast	59343798
53	male	28,88	0	no	northwest	98698102
18	male	41,14	0	no	southeast	11467966
51	male	35,97	1	no	southeast	93861613
31	female	29,26	1	no	southeast	43505144
60	male	36955	0	no	northeast	1,27E+09
21	male	36,86	0	no	northwest	19173184
29	male	22515	3	no	northeast	5,21E+08
62	female	29,92	0	no	southeast	1,35E+08
39	female	41,8	0	no	southeast	5662225
22	female	23,18	0	no	northeast	27319122
39	female	31,92	2	no	northwest	72094918
30	male	44,22	2	no	southeast	42661658
30	female	22895	1	no	northeast	4,72E+08
58	female	33,1	0	no	southwest	11848141
33	male	24795	0	yes	northeast	1,79E+09
42	female	26,18	1	no	southeast	70467222
64	female	35,97	0	no	southeast	1,43E+08
18	female	42,24	0	yes	southeast	3,88E+08
23	male	26,51	0	no	southeast	18158759
45	female	35815	0	no	northwest	7,73E+08
40	female	41,42	1	no	northwest	2,85E+09
19	female	36575	0	no	northwest	2,14E+08
18	male	30,14	0	no	southeast	11315066
25	male	25,84	1	no	northeast	33097926
33	female	42,94	3	no	northwest	63609936
28	male	22515	2	no	northeast	4,43E+08
36	male	34,43	2	no	southeast	55843057
20	female	31,46	0	no	southeast	18779294
24	female	24225	0	no	northwest	2,84E+08
23	male	37,1	3	no	southwest	3597596
47	female	26125	1	yes	northeast	2,34E+09
33	female	35,53	0	yes	northwest	5,51E+09
45	male	33,7	1	no	southwest	7445918
26	male	17,67	0	no	northwest	26809493
18	female	31,13	0	no	southeast	16218827
44	female	29,81	2	no	southeast	82192039
60	male	24,32	0	no	northwest	1,25E+08
64	female	31825	2	no	northeast	1,61E+09
56	male	31,79	2	yes	southeast	4,38E+08
36	male	28025	1	yes	northeast	2,08E+09

41	male	30,78	3	yes	northeast	3,96E+08
39	male	21,85	1	no	northwest	61174945
63	male	33,1	0	no	southwest	13393756
36	female	25,84	0	no	northwest	52663656
28	female	23845	2	no	northwest	4,72E+08
58	male	34,39	0	no	northwest	1,17E+08
36	male	33,82	1	no	northwest	53774578
42	male	35,97	2	no	southeast	71603303
56	female	28,31	0	no	northeast	1,17E+08
35	female	23465	2	no	northeast	6,4E+08
59	female	31,35	0	no	northwest	1,26E+08
23	female	32,78	2	yes	southeast	3,6E+08
57	female	29,81	0	yes	southeast	2,75E+08
53	male	30495	0	no	northeast	1,01E+09
60	female	32,45	0	yes	southeast	4,5E+08
51	female	34,2	1	no	southwest	9872701
23	male	50,38	1	no	southeast	24380552
55	male	32775	0	no	northwest	1,06E+09
37	female	30,78	0	yes	northeast	3,73E+08
61	male	32,3	2	no	northwest	14119,62
46	female	35,53	0	yes	northeast	4,21E+08
53	female	23,75	2	no	northeast	1,17E+08
49	female	23845	3	yes	northeast	2,41E+09
48	female	33,11	0	yes	southeast	4,1E+08
25	male	24,13	0	yes	northwest	1,58E+08
25	female	32,23	1	no	southeast	1,82E+09
37	female	47,6	2	yes	southwest	46113511
38	female	28	3	no	southwest	7151092
55	female	33535	2	no	northwest	1,23E+09
36	female	19855	0	no	northeast	5,46E+08
18	male	37,29	0	no	southeast	11414451
57	male	43,7	1	no	southwest	11576,13
61	male	23655	0	no	northeast	1,31E+09
50	male	36,2	0	no	southwest	8457818
26	female	29,48	1	no	southeast	33923652
42	male	24,86	0	no	southeast	59668874
44	male	21,85	3	no	northeast	88911395
33	male	33,44	5	no	southeast	66537886
22	male	34,8	3	no	southwest	3443064
23	male	27,36	1	no	northwest	27890574
21	female	22135	0	no	northeast	2,59E+08
51	female	37,05	3	yes	northeast	4,63E+08
25	male	26695	4	no	northwest	4,88E+08
32	male	28,93	1	yes	southeast	1,97E+08
57	male	28975	0	yes	northeast	2,72E+09
22	male	39,5	0	no	southwest	1682597
57	male	33,63	1	no	northwest	1,19E+08
64	female	26885	0	yes	northwest	2,93E+09

54	male	24035	0	no	northeast	1,04E+09
47	male	38,94	2	yes	southeast	4,42E+08
62	male	32,11	0	no	northeast	1,36E+08
61	female	44	0	no	southwest	13063883
43	female	20045	2	yes	northeast	1,98E+09
19	male	25555	1	no	northwest	2,22E+08
18	female	40,26	0	no	southeast	16345734
19	female	22515	0	no	northwest	2,12E+08
49	male	22515	0	no	northeast	8,69E+08
60	male	40,92	0	yes	southeast	4,87E+08
26	male	27265	3	no	northeast	4,66E+08
49	male	36,85	0	no	southeast	81257845
60	female	35,1	0	no	southwest	12644589
26	female	29355	2	no	northeast	4,56E+08
27	male	32585	3	no	northeast	4,85E+08
44	female	32,34	1	no	southeast	76337206
63	male	39,8	3	no	southwest	15170069
22	male	28,31	1	no	northwest	26390429
18	male	31,73	0	yes	northeast	3,37E+08
59	female	26695	3	no	northwest	1,44E+09
33	male	24605	2	no	northwest	5,26E+08
24	female	33,99	0	no	southeast	24733341
43	female	26885	0	yes	northwest	2,18E+09
45	male	22895	0	yes	northeast	3,51E+09
35	female	34,21	1	no	southeast	52452269
62	female	25	0	no	southwest	13451122
62	female	33,2	0	no	southwest	13462,52
38	male	31	1	no	southwest	5488262
34	male	35815	0	no	northwest	4,32E+08
50	male	32,11	2	no	northeast	2,53E+09
62	female	39,16	0	no	southeast	1,35E+08
41	male	34,21	1	no	southeast	62897549
26	male	46,53	1	no	southeast	29270647
39	female	32,5	1	no	southwest	6238298
45	female	35,3	0	no	southwest	7348142
32	male	37,18	2	no	southeast	46733922
44	male	29735	2	no	northeast	3,21E+09
39	female	24225	5	no	northwest	8,97E+08
18	male	26,18	2	no	southeast	23040022
53	male	29,48	0	no	southeast	94876442
18	male	23,21	0	no	southeast	11218739
50	female	46,09	1	no	southeast	95495651
18	female	40185	0	no	northeast	2,22E+08
19	male	22,61	0	no	northwest	16284709
62	male	39,93	0	no	southeast	1,3E+08
56	female	35,8	1	no	southwest	11674,13
42	male	35,8	2	no	southwest	7160094
37	male	34,2	1	yes	northeast	39047285

42	male	31255	0	no	northwest	6,36E+08
57	male	18335	0	no	northeast	1,15E+09
51	male	42,9	2	yes	southeast	47462894
30	female	28405	1	no	northwest	4,53E+08
44	male	30,2	2	yes	southwest	38998546
34	male	27835	1	yes	northwest	2E+09
31	male	39,49	1	no	southeast	38757341
24	male	26,79	1	no	northwest	1,26E+09
43	male	34,96	1	yes	northeast	4,1E+08
48	male	36,67	1	no	northwest	2,85E+09
19	female	39615	1	no	northwest	2,73E+08
63	female	35,2	1	no	southeast	14474675
46	male	24795	3	no	northeast	9,5E+08
52	male	36765	2	no	northwest	2,65E+09
51	male	24795	2	yes	northwest	2,4E+09
44	male	25365	1	no	northwest	7,52E+08
21	male	25745	2	no	northeast	3,28E+08
39	female	34,32	5	no	southeast	85968278
50	female	28,16	3	no	southeast	1,07E+08
34	female	23,56	0	no	northeast	49923764
22	female	20235	0	no	northwest	2,53E+08
19	female	40,5	0	no	southwest	1759338
26	male	35,42	0	no	southeast	23226218
29	male	22895	0	yes	northeast	1,61E+09
48	male	40,15	0	no	southeast	78041605
26	male	29,15	1	no	southeast	29029065
45	female	39995	3	no	northeast	9,7E+08
36	female	29,92	0	no	southeast	48890368
54	male	25,46	1	no	northeast	2,55E+09
34	male	21375	0	no	northeast	4,5E+08
27	female	30,59	1	no	northeast	1,68E+09
20	male	30115	5	no	northeast	4,92E+08
43	male	30115	3	no	northwest	8,41E+08
45	female	27645	1	no	northwest	2,83E+09
34	male	34675	0	no	northeast	4,52E+08
24	female	20,52	0	yes	northeast	1,46E+08
38	female	27835	2	no	northeast	7,14E+08
50	female	31,6	2	no	southwest	10118424
38	male	28,27	1	no	southeast	54844673
27	female	20045	3	yes	northwest	1,64E+09
39	female	23275	3	no	northeast	7,99E+08
39	female	34,1	3	no	southwest	7418522
63	female	36,85	0	no	southeast	1,39E+08
33	female	36,29	3	no	northeast	65517501
36	female	26885	0	no	northwest	5,27E+08
30	male	22,99	2	yes	northwest	1,74E+08
24	male	32,7	0	yes	southwest	34472841
47	male	19,19	1	no	northeast	86275411

29	male	31,73	2	no	northwest	44333877
28	male	29,26	2	no	northeast	44382634
47	male	28215	3	yes	northwest	2,49E+09
25	male	24985	2	no	northeast	2,32E+09
51	male	27,74	1	no	northeast	99577216
43	male	20,13	2	yes	southeast	1,88E+08
61	female	33,33	4	no	southeast	3,66E+09
48	male	32,3	1	no	northwest	8765249
59	male	25,46	0	no	northwest	1,21E+08
19	female	24605	1	no	northwest	2,71E+08
26	female	34,2	2	no	southwest	3987926
54	female	35815	3	no	northwest	1,25E+09
21	female	32,68	2	no	northwest	2,6E+09
51	male	37	0	no	southwest	8798593
22	female	31,02	3	yes	southeast	3,56E+08
47	male	36,08	1	yes	southeast	4,22E+08
18	male	23,32	1	no	southeast	17110268
47	female	45,32	1	no	southeast	85698618
21	female	34,6	0	no	southwest	2020177
23	male	18715	0	no	northwest	2,16E+09
54	male	31,6	0	no	southwest	9850432
37	female	17,29	2	no	northeast	68779801
46	female	23655	1	yes	northwest	2,17E+09
55	female	35,2	0	yes	southeast	44423803
30	female	27,93	0	no	northeast	41375227
18	male	21565	0	yes	northeast	1,37E+09
61	male	38,38	0	no	northwest	1,3E+08
54	female	23	3	no	southwest	12094478
22	male	37,07	2	yes	southeast	3,75E+08
45	female	30495	1	yes	northwest	3,97E+09
22	male	28,88	0	no	northeast	22508352
19	male	27265	2	no	northwest	2,25E+09
35	female	28025	0	yes	northwest	2,02E+09
18	male	23085	0	no	northeast	1,7E+08
20	male	30685	0	yes	northeast	3,35E+09
55	male	35245	1	no	northeast	1,14E+09
22	male	52,58	1	yes	southeast	4,45E+08
25	female	22515	1	no	northwest	3,59E+08
44	female	36955	1	no	northwest	8,02E+08
64	male	26,41	0	no	northeast	1,44E+08
49	male	29,83	1	no	northeast	92880267
27	female	21,47	0	no	northwest	33534703
55	male	27645	0	no	northwest	1,06E+09
45	female	31,79	0	no	southeast	1,79E+09
24	female	39,49	0	no	southeast	24809791
32	male	33,82	1	no	northwest	44627218
24	male	32,01	0	no	southeast	19815819
57	male	27,94	1	no	southeast	1,16E+08

59	male	41,14	1	yes	southeast	4,9E+08
36	male	28595	3	no	northwest	6,55E+08
48	male	37,29	2	no	southeast	89781851
39	male	42655	0	no	northeast	5,76E+08
63	male	21,66	1	no	northwest	1,43E+08
54	female	31,9	1	no	southeast	10928849
37	male	37,07	1	yes	southeast	3,99E+08
63	male	31445	0	no	northeast	1,4E+09
21	male	31255	0	no	northwest	1,91E+08
54	female	28,88	2	no	northeast	1,21E+08
60	female	18335	0	no	northeast	1,32E+09
32	female	29,59	1	no	southeast	45628421
47	female	32	1	no	southwest	8551347
28	male	31,68	0	yes	southeast	3,47E+08
63	male	33,66	3	no	southeast	1,52E+08
18	male	21,78	2	no	southeast	1,19E+09
32	male	27835	1	no	northwest	4,45E+08
38	male	19,95	1	no	northwest	58559025
62	female	30495	2	no	northwest	1,5E+09
55	male	28975	0	no	northeast	1,08E+09
57	male	31,54	0	no	northwest	1,14E+08
52	male	47,74	1	no	southeast	97489106
47	male	36,19	0	yes	southeast	4,17E+08
55	female	29,83	0	no	northeast	1,13E+08
23	male	32,7	3	no	southwest	3591,48
50	female	33,7	4	no	southwest	11299343
18	female	31,35	4	no	northeast	45611885
51	female	34,96	2	yes	northeast	4,46E+08
22	male	33,77	0	no	southeast	16746323
52	female	30875	0	no	northeast	2,3E+09
25	female	33,99	1	no	southeast	32271211
33	female	19095	2	yes	northeast	1,68E+09
29	male	38,94	1	no	southeast	34714096
58	male	36,08	0	no	southeast	1,14E+08
54	female	31,24	0	no	southeast	1,03E+08
49	female	29925	0	no	northwest	8,99E+08
50	female	26,22	2	no	northwest	1,05E+08
26	male	30	1	no	southwest	2904088
45	male	20,35	3	no	southeast	86053615
54	female	32,3	1	no	northeast	11512405
38	male	38,39	3	yes	southeast	4,19E+08
48	female	25,85	3	yes	southeast	2,42E+08
28	female	26315	3	no	northwest	5,31E+08
23	male	24,51	0	no	northeast	23960959
55	male	32,67	1	no	southeast	1,08E+08
41	male	29,64	5	no	northeast	92224026
25	male	33,33	2	yes	southeast	3,61E+08
33	male	35,75	1	yes	southeast	3,83E+08

30	female	19,95	3	no	northwest	56934305
23	female	31,4	0	yes	southwest	34166273
46	male	38,17	2	no	southeast	83471643
53	female	36,86	3	yes	northwest	4,67E+08
27	female	32395	1	no	northeast	1,89E+09
23	female	42,75	1	yes	northeast	4,09E+08
35	female	35,86	2	no	southeast	58365204
34	male	32,8	1	no	southwest	1,44E+09
39	female	23,87	5	no	southeast	85823023
27	male	45,9	2	no	southwest	3693428
57	male	40,28	0	no	northeast	2,07E+09
52	female	18335	0	no	northwest	9,99E+08
28	male	33,82	0	no	northwest	1,97E+09
44	female	25	1	no	southwest	7623518
26	female	22,23	0	no	northwest	31762877
33	male	30,25	0	no	southeast	37043545
19	female	32,49	0	yes	northwest	3,69E+09
50	male	37,07	1	no	southeast	90480273
41	female	32,6	3	no	southwest	7954517
52	female	24,86	0	no	southeast	2,71E+09
39	male	32,34	2	no	southeast	63380756
50	male	32,3	2	no	southwest	9630397
52	male	32775	3	no	northwest	1,13E+09
60	male	32,8	0	yes	southwest	5,26E+09
20	female	31,92	0	no	northwest	22615688
42	male	34,1	0	no	southwest	5979731
18	female	30305	0	no	northeast	2,2E+08
58	female	36,48	0	no	northwest	1,22E+08
43	female	32,56	3	yes	southeast	4,09E+08
35	female	35815	1	no	northwest	5,63E+08
48	female	27,93	4	no	northwest	1,1E+08
36	female	22135	3	no	northeast	7,23E+08
19	male	44,88	0	yes	southeast	3,97E+08
23	female	23,18	2	no	northwest	1,44E+09
20	female	30,59	0	no	northeast	24597201
32	female	41,1	0	no	southwest	3989841
43	female	34,58	1	no	northwest	77272532
34	male	42,13	2	no	southeast	51241887
30	male	38,83	1	no	southeast	1,9E+09
18	female	28215	0	no	northeast	2,2E+08
41	female	28,31	1	no	northwest	71535539
35	female	26125	0	no	northeast	5,23E+08
57	male	40,37	0	no	southeast	1,1E+08
32	male	35,2	2	no	southwest	4670,64
37	female	34105	1	no	northwest	6,11E+08
18	male	27,36	1	yes	northeast	1,72E+08
56	female	41,91	0	no	southeast	1,11E+08
38	male	29,26	2	no	northwest	64578434

29	male	32,11	2	no	northwest	44339159
52	female	24,13	1	yes	northwest	2,39E+08
23	female	34865	0	no	northeast	2,9E+08
31	male	29,81	0	yes	southeast	1,94E+08
42	female	41325	1	no	northeast	7,65E+08
24	female	29925	0	no	northwest	2,85E+08
48	female	27,36	1	no	northeast	94473824
23	female	28,49	1	yes	southeast	1,83E+08
45	male	23,56	2	no	northeast	86038234
20	male	35625	3	yes	northwest	3,75E+09
62	female	32,68	0	no	northwest	1,38E+08
43	female	25,27	1	yes	northeast	2,18E+08
23	female	28	0	no	southwest	1,31E+09
31	female	32775	2	no	northwest	5,33E+08
41	female	21755	1	no	northeast	1,37E+09
58	female	32395	1	no	northeast	1,3E+09
48	female	36575	0	no	northwest	8,67E+08
31	female	21755	0	no	northwest	4,13E+08
19	female	27,93	3	no	northwest	1,88E+09
19	female	30,02	0	yes	northwest	3,33E+08
41	male	33,55	0	no	southeast	56998375
40	male	29355	1	no	northwest	6,39E+08
37	male	24,32	2	no	northwest	61987518
46	male	40375	2	no	northwest	8,73E+08
22	male	32,11	0	no	northwest	20553249
18	female	27,28	3	yes	southeast	1,82E+08
35	male	17,86	1	no	northwest	51165004
59	female	34,8	2	no	southwest	3,69E+09
36	male	33,4	2	yes	southwest	38415474
37	female	25555	1	yes	northeast	2,03E+09
59	male	37,1	1	no	southwest	12347172
36	male	30875	1	no	northwest	5,37E+08
39	male	34,1	2	no	southeast	2,36E+09
18	male	21,47	0	no	northeast	17024553
52	female	33,3	2	no	southwest	10806839
27	female	31255	1	no	northwest	3,96E+08
18	male	39,14	0	no	northeast	1,29E+09
29	male	37,29	2	no	southeast	40581161
46	female	34,6	1	yes	southwest	41661602
38	female	30,21	3	no	northwest	75371639
30	female	21945	1	no	northeast	4,72E+08
40	male	24,97	2	no	southeast	65935083
20	female	24,42	0	yes	southeast	2,61E+09
41	male	23,94	1	no	northeast	68584796
33	female	39,82	1	no	southeast	47956568
38	male	16815	2	no	northeast	6,64E+08
42	male	37,18	2	no	southeast	71620122
56	male	34,43	0	no	southeast	1,06E+08

58	male	30305	0	no	northeast	1,19E+09
52	male	34485	3	yes	northwest	6E+09
54	female	24605	3	no	northwest	1,25E+09
45	female	27,83	2	no	southeast	85157587
26	male	31065	0	no	northwest	2,7E+08
63	female	21,66	0	no	northeast	1,44E+08
58	female	28215	0	no	northwest	1,22E+09
37	male	22705	3	no	northeast	6,99E+08
25	female	42,13	1	no	southeast	32384357
52	male	41,8	2	yes	southeast	47269854
64	male	36,96	2	yes	southeast	4,96E+08
22	female	21,28	3	no	northwest	42962712
28	female	33,11	0	no	southeast	31716149
18	male	33,33	0	no	southeast	11359407
18	female	39,82	0	no	southeast	16339618
32	male	33,63	1	yes	northeast	3,76E+08
24	male	29,83	0	yes	northeast	1,86E+08
34	female	27,72	0	no	southeast	44151588
42	female	37,9	0	no	southwest	6474013
51	female	36385	3	no	northwest	1,14E+09
54	female	27645	1	no	northwest	1,13E+09
55	male	37715	3	no	northwest	3,01E+09
52	female	23,18	0	no	northeast	1,02E+08
32	female	20,52	0	no	northeast	45442348
28	male	37,1	1	no	southwest	3277161
49	female	33345	2	no	northeast	1,04E+09
64	male	23,76	0	yes	southeast	2,69E+08
24	male	31065	0	yes	northeast	3,43E+09
20	female	33,3	0	no	southwest	1880487
26	male	33915	1	no	northwest	3,29E+08
25	female	34485	0	no	northwest	3,02E+08
43	male	25,52	5	no	southeast	1,45E+09
35	male	27,61	1	no	southeast	47470529
32	female	29735	0	no	northwest	4,36E+08
39	male	29925	1	yes	northeast	2,25E+09
25	female	26,79	2	no	northwest	41891131
48	female	33,33	0	no	southeast	82836807
47	female	27645	2	yes	northwest	2,45E+09
18	female	21,66	0	yes	northeast	1,43E+08
61	male	36,3	1	yes	southwest	47403,88
47	female	24,32	0	no	northeast	85346718
28	female	17,29	0	no	northeast	37326251
20	male	39,4	2	yes	southwest	38344566
44	male	34,32	1	no	southeast	71474728
38	female	19,95	2	no	northeast	71339025
19	male	34,9	0	yes	southwest	34828654
21	male	23,21	0	no	southeast	15153449
46	male	25745	3	no	northwest	9,3E+08

58	male	25175	0	no	northeast	1,19E+09
20	male	22	1	no	southwest	1964,78
18	male	26125	0	no	northeast	1,71E+08
28	female	26,51	2	no	southeast	43404409
33	male	27455	2	no	northwest	5,26E+08
19	female	25745	1	no	northwest	2,71E+08
45	male	30,36	0	yes	southeast	6,26E+09
62	male	30875	3	yes	northwest	4,67E+09
42	male	24605	2	yes	northeast	2,13E+09
24	female	27,72	0	no	southeast	24646188
29	female	21,85	0	yes	northeast	1,61E+08
25	female	30,2	0	yes	southwest	33900653
41	male	32,2	2	no	southwest	6875961
42	male	26315	1	no	northwest	6,94E+08
33	female	26695	0	no	northwest	4,57E+08
34	male	42,9	1	no	southwest	4536259
19	female	34,7	2	yes	southwest	36397576
30	female	23655	3	yes	northwest	1,88E+09
18	male	28,31	1	no	northeast	1,13E+09
18	male	53,13	0	no	southeast	11634627
35	male	39,71	4	no	northeast	1,95E+09
39	female	26315	2	no	northwest	7,2E+08
31	male	31065	3	no	northwest	5,43E+08
62	male	26695	0	yes	northeast	2,81E+09
62	male	38,83	0	no	southeast	1,3E+08
42	female	40,37	2	yes	southeast	4,39E+08
31	male	25935	1	no	northwest	4,24E+08
61	male	33535	0	no	northeast	1,31E+09
42	female	32,87	0	no	northeast	70500213
23	female	24225	2	no	northeast	2,24E+09
52	male	38,6	2	no	southwest	10325206
57	female	25,74	2	no	southeast	1,26E+08
23	female	33,4	0	no	southwest	1,08E+09
52	female	44,7	3	no	southwest	11411685
50	male	30,97	3	no	northwest	1,06E+08
18	female	31,92	0	no	northeast	22059808
18	female	36,85	0	no	southeast	16298335

3. Analisis seluruh variable data dengan statistik dekriptif, uji normalitas distribusi data dan berikan visualisasi datanya (histogram & box plot) serta berikan kesimpulan dari hasil analisis tersebut!

Jawab

1. **Ringkasan statistik deskriptif**

a. Umur: rata-rata sekitar 39 sampai 40 tahun, median 40 tahun, rentang 18 sampai 64 tahun. Sebaran cukup merata dari usia muda hingga usia lanjut.
b. Indeks massa tubuh: median sekitar 35.6, namun terdapat nilai sangat besar yang membuat nilai rata-rata tampak tidak realistis. Ini menandakan ada nilai ekstrem yang kuat sehingga median lebih mewakili pusat data.
c. Jumlah anak: rata-rata sekitar satu anak, median satu anak, rentang 0 sampai 5 anak. Sebaran condong ke kanan karena lebih banyak yang memiliki sedikit anak.
d. Biaya asuransi: median sekitar seratus dua puluh enam juta, dengan rentang sangat lebar. Nilai ekstrem tinggi cukup banyak sehingga sebaran sangat melebar dan rata-rata jauh lebih besar daripada median.

2. **Uji normalitas**
   a. Hasil uji Shapiro Wilk menunjukkan semua variabel numerik tidak mengikuti distribusi normal. Ini sejalan dengan pengamatan deskriptif bahwa beberapa variabel memiliki sebaran miring dan banyak nilai ekstrem.
3. **Gambaran histogram dan box plot**
   a. Histogram memperlihatkan sebaran tidak simetris pada biaya asuransi dan indeks massa tubuh. Puncak sebaran tidak tunggal dan ekor panjang ke kanan.
   b. Box plot menegaskan keberadaan pencilan pada biaya asuransi dan indeks massa tubuh. Pada umur dan jumlah anak pencilan relatif lebih sedikit, tetapi jumlah anak tetap tampak condong ke kanan.
4. **Kesimpulan analisis**
   Seluruh variabel numerik tidak berdistribusi normal dan memiliki pencilan, terutama pada biaya asuransi dan indeks massa tubuh. Untuk pemusatan, median dan rentang antarkuartil lebih mewakili daripada rata-rata dan simpangan baku. Jika akan dilakukan pemodelan atau uji yang sensitif terhadap normalitas, pertimbangkan transformasi data (misalnya log pada biaya) atau gunakan metode yang tahan terhadap pencilan dan tidak mensyaratkan normalitas. Visualisasi histogram dan box plot mendukung

4. Analisis Korelasi dan Regresi antra lama belajar dengan hasil skor, berikan kesimpulan atas analisis tersebut!

Jawab

1. Korelasi usia dengan charges

- **Nilai korelasi r = 0,1451**

- **Artinya:**
    - **Hubungan positif tapi lemah**
    - **Semakin bertambah usia, biaya asuransi (charges) cenderung ikut naik**
    - **Tapi pengaruhnya tidak terlalu kuat**

Jadi: usia berpengaruh, tapi hanya sedikit terhadap besar-kecilnya biaya asuransi.

---

2. Regresi usia dengan charges

Persamaan regresi:

Charges = 173.093.504,84 + 10.327.166,00 × Age

Artinya secara sederhana:

- **Setiap usia naik 1 tahun, biaya asuransi diprediksi naik sekitar 10,3 juta (sesuai skala datanya).**
- **Nilai R^2 = 0,0211 (2,11%)**
  → Hanya 2,11% variasi charges yang bisa dijelaskan oleh usia.
  → Hampir semua (97,89%) dipengaruhi faktor lain (smoker, BMI, dll).

---

3. Kesimpulan

- **Usia dan charges punya hubungan positif tapi lemah.**
- **Usia bukan faktor utama penentu biaya asuransi.**
- **Untuk analisis/prediksi yang lebih baik, perlu menambah variabel lain (smoker, BMI, children, region, dll), bukan hanya umur saja.**

5. Tuliskan kodingan Python dari awal input data sampai analisis korelasi disertai komentar untuk apa kodingan-kodingan tersebut!

Jawab

```
import warnings

# Import library yang diperlukan untuk analisis data
import matplotlib.pyplot as plt  # Untuk visualisasi grafik
import numpy as np  # Untuk operasi numerik
import pandas as pd  # Untuk manipulasi dan analisis data
```

```
import scipy.stats as stats  # Untuk uji statistik
import seaborn as sns  # Untuk visualisasi yang lebih menarik
import statsmodels.formula.api as smf  # Untuk analisis regresi

# Ignore warnings untuk output yang lebih bersih
warnings.filterwarnings('ignore')

# Set style untuk visualisasi agar lebih menarik
sns.set_style("whitegrid")
plt.rcParams['figure.figsize'] = (12, 6)

#
==========================================================================
==
# 1. MEMUAT DATA (DATA LOADING)
#
==========================================================================
==
# Tahap pertama adalah memuat data dari file Excel ke dalam DataFrame pandas

print("=" * 80)
print("1. MEMUAT DATA DARI FILE")
print("=" * 80)

# Membaca file Excel menggunakan pandas
df = pd.read_excel("./insurance_remed.xlsx")

print(f"✓ Data berhasil dimuat dari file 'insurance_remed.xlsx'")
print(f"✓ Jumlah data yang dimuat: {df.shape[0]} baris x {df.shape[1]}
kolom")
print(f"\n5 Baris Pertama Data:")
print(df.head())

#
==========================================================================
==
# 2. ANALISIS DATA AWAL (INITIAL DATA ANALYSIS)
#
==========================================================================
==
# Melakukan pengecekan menyeluruh pada dataset untuk mengetahui:
# - Ukuran dataset (jumlah baris dan kolom)
# - Tipe data dari setiap variabel
# - Keberadaan missing values (nilai yang hilang)
# - Keberadaan data duplikat
# - Keberadaan outlier (data pencilan)

print("\n" + "=" * 80)
```

```
print("2. ANALISIS DATA AWAL")
print("=" * 80)

# A. UKURAN DATASET
print("\n[A] UKURAN DATASET")
print("-" * 80)
print(f"Jumlah baris (observasi): {df.shape[0]}")
print(f"Jumlah kolom (variabel): {df.shape[1]}")
print(f"Total ukuran dataset: {df.shape[0]} baris x {df.shape[1]} kolom")

# B. TIPE DATA SETIAP VARIABEL
print("\n[B] TIPE DATA SETIAP VARIABEL")
print("-" * 80)
print("\nInformasi Detail Dataset:")
df.info()
print("\nRingkasan Tipe Data:")
for col in df.columns:
    print(f"  - {col:20s}: {df[col].dtype}")

# C. MISSING VALUES (Nilai yang Hilang)
print("\n[C] MISSING VALUES (Nilai yang Hilang)")
print("-" * 80)
missing_values = df.isnull().sum()
missing_percentage = (df.isnull().sum() / len(df)) * 100
missing_df = pd.DataFrame({
    'Kolom': missing_values.index,
    'Jumlah Missing': missing_values.values,
    'Persentase (%)': missing_percentage.values
})
print(missing_df.to_string(index=False))

total_missing = df.isnull().sum().sum()
if total_missing > 0:
    print(f"\n⚠ Total missing values: {total_missing} ({(total_missing /
(df.shape[0] * df.shape[1])) * 100:.2f}%)")
else:
    print(f"\n✓ Tidak ada missing values dalam dataset")

# D. DATA DUPLIKAT
print("\n[D] DATA DUPLIKAT")
print("-" * 80)
duplicate_rows = df.duplicated().sum()
print(f"Jumlah baris duplikat: {duplicate_rows}")
if duplicate_rows > 0:
    print(f"⚠ Terdapat {duplicate_rows} baris duplikat yang perlu dihapus")
    print(f"Persentase duplikasi: {(duplicate_rows / len(df)) * 100:.2f}%")
else:
    print("✓ Tidak ada data duplikat")
```

```
# E. TIPE DATA YANG SALAH
print("\nKonversi Tipe Data")
print("-" * 80)
df['bmi'] = pd.to_numeric(df['bmi'], errors='coerce')
df['charges'] = pd.to_numeric(df['charges'], errors='coerce')

print("✓ Kolom 'bmi' dikonversi ke tipe numerik")
print("✓ Kolom 'charges' dikonversi ke tipe numerik")

# F. DETEKSI OUTLIER (Data Pencilan)
print("\n[F] DETEKSI OUTLIER (Data Pencilan)")
print("-" * 80)

# Konversi kolom numerik untuk deteksi outlier
df_numeric_check = df.copy()
df_numeric_check['age_numeric'] = pd.to_numeric(df_numeric_check['age'],
errors='coerce')
df_numeric_check['bmi_numeric'] = pd.to_numeric(df_numeric_check['bmi'],
errors='coerce')
df_numeric_check['charges_numeric'] =
pd.to_numeric(df_numeric_check['charges'], errors='coerce')

# Visualisasi Boxplot untuk deteksi outlier
fig, axes = plt.subplots(1, 3, figsize=(15, 5))

# Boxplot untuk Age
axes[0].boxplot(df_numeric_check['age_numeric'].dropna(), vert=True)
axes[0].set_title('Boxplot: Age (Umur)', fontsize=12, fontweight='bold')
axes[0].set_ylabel('Umur (tahun)')
axes[0].grid(True, alpha=0.3)

# Boxplot untuk BMI
axes[1].boxplot(df_numeric_check['bmi_numeric'].dropna(), vert=True)
axes[1].set_title('Boxplot: BMI (Body Mass Index)', fontsize=12,
fontweight='bold')
axes[1].set_ylabel('BMI')
axes[1].grid(True, alpha=0.3)

# Boxplot untuk Charges
axes[2].boxplot(df_numeric_check['charges_numeric'].dropna(), vert=True)
axes[2].set_title('Boxplot: Charges (Biaya Asuransi)', fontsize=12,
fontweight='bold')
axes[2].set_ylabel('Biaya (dollar)')
axes[2].grid(True, alpha=0.3)

plt.tight_layout()
plt.savefig('boxplot_outlier_detection.png', dpi=300, bbox_inches='tight')
```

```
plt.close()

print("✓ Boxplot telah dibuat untuk mendeteksi outlier pada variabel
numerik")
print("  (Outlier ditandai dengan titik-titik di luar whisker)")

#
==========================================================================
==
# KESIMPULAN ANALISIS DATA AWAL
#
==========================================================================
==

print("\n" + "=" * 80)
print("KESIMPULAN ANALISIS DATA AWAL")
print("=" * 80)

print(f"""
RINGKASAN TEMUAN:

1. UKURAN DATASET:
   - Dataset memiliki {df.shape[0]} baris (observasi) dan {df.shape[1]} kolom
(variabel)
   - Ukuran dataset cukup besar untuk analisis statistik

2. TIPE DATA:
   - age: integer (sudah benar)
   - sex: object/string (sudah benar untuk data kategorikal)
   - bmi: object (BERMASALAH - seharusnya numerik)
   - children: integer (sudah benar)
   - smoker: object/string (sudah benar untuk data kategorikal)
   - region: object/string (sudah benar untuk data kategorikal)
   - charges: object (BERMASALAH - seharusnya numerik)

3. MISSING VALUES:
   - bmi: {missing_values['bmi']} nilai hilang
   - charges: {missing_values['charges']} nilai hilang
   - Variabel lain: tidak ada missing values

4. DATA DUPLIKAT:
   - Terdapat {duplicate_rows} baris duplikat yang perlu dihapus

5. TIPE DATA YANG SALAH:
   - Kolom 'bmi' berisi nilai non-numerik (tanggal/teks)
   - Kolom 'charges' berisi nilai non-numerik

6. OUTLIER:
```

```
   - Terdeteksi outlier pada beberapa variabel (lihat boxplot)
   - Outlier pada 'charges' mungkin normal (biaya tinggi untuk kondisi
tertentu)

⚠ KESIMPULAN AKHIR:
DATA INI TIDAK DAPAT LANGSUNG DIOLAH karena memiliki beberapa masalah:
  ✗ Ada missing values pada kolom bmi dan charges
  ✗ Ada data duplikat yang perlu dihapus
  ✗ Ada tipe data yang salah (bmi dan charges berisi nilai non-numerik)
  ✗ Perlu pembersihan data terlebih dahulu sebelum analisis lanjutan

SOLUSI:
  → Hapus data duplikat
  → Hapus baris dengan tipe data yang salah
  → Tangani missing values (hapus atau imputasi)
  → Konversi tipe data ke format yang benar
""")

#
================================================================================
==
# 3. PEMBERSIHAN DATA (DATA CLEANING)
#
================================================================================
==
# Melakukan pembersihan data dengan:
# - Menghapus data duplikat
# - Menghapus data dengan tipe yang salah (non-numerik pada kolom numerik)
# - Menangani missing values (dengan menghapus baris yang memiliki missing
values)

print("\n" + "=" * 80)
print("3. PEMBERSIHAN DATA (DATA CLEANING)")
print("=" * 80)

# Menyimpan ukuran data original
original_size = df.shape[0]
print(f"\nUkuran data SEBELUM pembersihan: {original_size} baris")

# LANGKAH 1: Menghapus data duplikat
print("\n[LANGKAH 1] Menghapus Data Duplikat")
print("-" * 80)
duplicates_count = df.duplicated().sum()
df = df.drop_duplicates()
print(f"✓ Dihapus: {duplicates_count} baris duplikat")
print(f"✓ Sisa data: {df.shape[0]} baris")

# LANGKAH 2: Menghapus baris dengan missing values
```

```
print("\n[LANGKAH 2] Menghapus Missing Values")
print("-" * 80)
before_dropna = df.shape[0]
df_clean = df.dropna()
dropped_na = before_dropna - df_clean.shape[0]
print(f"✓ Dihapus: {dropped_na} baris dengan missing values")
print(f"✓ Sisa data: {df_clean.shape[0]} baris")

# LANGKAH 3: Menghapus baris dengan tipe data yang salah
print("\n[LANGKAH 3] Menghapus Data dengan Tipe yang Salah")
print("-" * 80)
before_type_filter = df_clean.shape[0]

# Filter hanya baris dengan BMI dan charges yang numerik
df_clean = df_clean[
    (df_clean['bmi'].apply(lambda x: isinstance(x, (int, float)))) &
    (df_clean['charges'].apply(lambda x: isinstance(x, (int, float))))
    ]

dropped_type = before_type_filter - df_clean.shape[0]
print(f"✓ Dihapus: {dropped_type} baris dengan tipe data yang salah")
print(f"✓ Sisa data: {df_clean.shape[0]} baris")

# LANGKAH 4: Konversi tipe data ke format yang benar
print("\n[LANGKAH 4] Konversi Tipe Data")
print("-" * 80)
df_clean['bmi'] = pd.to_numeric(df_clean['bmi'], errors='coerce')
df_clean['charges'] = pd.to_numeric(df_clean['charges'], errors='coerce')
print("✓ Kolom 'bmi' dikonversi ke tipe numerik")
print("✓ Kolom 'charges' dikonversi ke tipe numerik")

# LANGKAH 5: Menyimpan data bersih
print("\n[LANGKAH 5] Menyimpan Data Bersih")
print("-" * 80)
df_clean.to_excel("./insurance_remed_clean.xlsx", index=False)
print("✓ Data bersih disimpan ke file: insurance_remed_clean.xlsx")

# Ringkasan pembersihan
print("\n" + "=" * 80)
print("RINGKASAN PEMBERSIHAN DATA")
print("=" * 80)
print(f"Data SEBELUM pembersihan  : {original_size} baris")
print(f"Data SETELAH pembersihan  : {df_clean.shape[0]} baris")
print(f"Total data yang dihapus   : {original_size - df_clean.shape[0]}
baris")
print(f"Persentase data tersisa   : {(df_clean.shape[0] / original_size) *
100:.2f}%")
```

```
print(f"Persentase data dihapus   : {((original_size - df_clean.shape[0]) /
original_size) * 100:.2f}%")

print("\n✓ Data sudah bersih dan siap untuk analisis lanjutan!")

# Menampilkan info data bersih
print("\nInformasi Data Bersih:")
print("-" * 80)
df_clean.info()

#
==========================================================================
==
# 4. STATISTIK DESKRIPTIF (DESCRIPTIVE STATISTICS)
#
==========================================================================
==
# Menganalisis seluruh variabel dengan statistik deskriptif untuk memahami:
# - Ukuran pemusatan (mean, median)
# - Ukuran penyebaran (standar deviasi, range)
# - Nilai minimum dan maksimum

print("\n" + "=" * 80)
print("4. STATISTIK DESKRIPTIF")
print("=" * 80)

print("\n[A] STATISTIK DESKRIPTIF VARIABEL NUMERIK")
print("-" * 80)
desc_stats = df_clean.describe()
print(desc_stats)

print("\n[B] STATISTIK DESKRIPTIF VARIABEL KATEGORIKAL")
print("-" * 80)
print("\n1. Distribusi Jenis Kelamin (Sex):")
print(df_clean['sex'].value_counts())
print(f"   Persentase:")
print(df_clean['sex'].value_counts(normalize=True) * 100)

print("\n2. Distribusi Status Merokok (Smoker):")
print(df_clean['smoker'].value_counts())
print(f"   Persentase:")
print(df_clean['smoker'].value_counts(normalize=True) * 100)

print("\n3. Distribusi Wilayah (Region):")
print(df_clean['region'].value_counts())
print(f"   Persentase:")
print(df_clean['region'].value_counts(normalize=True) * 100)
```

```
print("\n[C] INTERPRETASI STATISTIK DESKRIPTIF")
print("-" * 80)
print(f"""
VARIABEL NUMERIK:

1. AGE (Umur):
   - Rata-rata: {desc_stats.loc['mean', 'age']:.2f} tahun
   - Median: {desc_stats.loc['50%', 'age']:.2f} tahun
   - Rentang: {desc_stats.loc['min', 'age']:.0f} - {desc_stats.loc['max',
'age']:.0f} tahun
   - Std Dev: {desc_stats.loc['std', 'age']:.2f} tahun
   → Usia pemegang polis bervariasi dari usia muda hingga lanjut

2. BMI (Body Mass Index):
   - Rata-rata: {desc_stats.loc['mean', 'bmi']:.2f}
   - Median: {desc_stats.loc['50%', 'bmi']:.2f}
   - Rentang: {desc_stats.loc['min', 'bmi']:.2f} - {desc_stats.loc['max',
'bmi']:.2f}
   - Std Dev: {desc_stats.loc['std', 'bmi']:.2f}
   → Mayoritas pemegang polis memiliki BMI normal hingga overweight

3. CHILDREN (Jumlah Anak):
   - Rata-rata: {desc_stats.loc['mean', 'children']:.2f} anak
   - Median: {desc_stats.loc['50%', 'children']:.0f} anak
   - Rentang: {desc_stats.loc['min', 'children']:.0f} - {desc_stats.loc['max',
'children']:.0f} anak
   → Sebagian besar pemegang polis memiliki sedikit tanggungan

4. CHARGES (Biaya Asuransi):
   - Rata-rata: ${desc_stats.loc['mean', 'charges']:,.2f}
   - Median: ${desc_stats.loc['50%', 'charges']:,.2f}
   - Rentang: ${desc_stats.loc['min', 'charges']:,.2f} -
${desc_stats.loc['max', 'charges']:,.2f}
   - Std Dev: ${desc_stats.loc['std', 'charges']:,.2f}
   → Biaya asuransi sangat bervariasi dengan standar deviasi yang tinggi
""")

#
================================================================================
==
# 5. UJI NORMALITAS DISTRIBUSI DATA
#
================================================================================
==
# Melakukan uji Shapiro-Wilk untuk menguji apakah data mengikuti distribusi
normal
# H0: Data mengikuti distribusi normal
# H1: Data tidak mengikuti distribusi normal
```

```
# Jika p-value > 0.05, maka gagal tolak H0 (data normal)
# Jika p-value <= 0.05, maka tolak H0 (data tidak normal)

print("\n" + "=" * 80)
print("5. UJI NORMALITAS DISTRIBUSI DATA (SHAPIRO-WILK TEST)")
print("=" * 80)

# Variabel numerik yang akan diuji
numeric_cols = ['age', 'bmi', 'children', 'charges']

# Melakukan uji Shapiro-Wilk untuk setiap variabel
print("\n[A] HASIL UJI SHAPIRO-WILK")
print("-" * 80)

shapiro_results = {}
for col in numeric_cols:
    # Ambil sample jika data terlalu besar (Shapiro-Wilk max 5000 data)
    if len(df_clean[col]) > 5000:
        sample_data = df_clean[col].sample(5000, random_state=42)
    else:
        sample_data = df_clean[col]

    stat, p_value = stats.shapiro(sample_data)
    shapiro_results[col] = {'statistic': stat, 'p_value': p_value}

    print(f"\nVariabel: {col.upper()}")
    print(f"  Statistik W: {stat:.6f}")
    print(f"  P-value: {p_value:.6f}")
    print(f"  Alpha (tingkat signifikansi): 0.05")

    if p_value > 0.05:
        print(f"  ✓ Kesimpulan: Data '{col}' MENGIKUTI distribusi normal")
        print(f"    (Gagal tolak H0 karena p-value > alpha)")
    else:
        print(f"  ✗ Kesimpulan: Data '{col}' TIDAK MENGIKUTI distribusi
normal")
        print(f"    (Tolak H0 karena p-value <= alpha)")

#
==========================================================================
==
# 6. VISUALISASI DISTRIBUSI DATA
#
==========================================================================
==
# Membuat histogram dan boxplot untuk setiap variabel numerik

print("\n" + "=" * 80)
```

```
print("6. VISUALISASI DISTRIBUSI DATA")
print("=" * 80)

# HISTOGRAM untuk setiap variabel numerik
print("\n[A] HISTOGRAM - Distribusi Frekuensi")
print("-" * 80)

fig, axes = plt.subplots(2, 2, figsize=(14, 10))
fig.suptitle('HISTOGRAM: Distribusi Frekuensi Variabel Numerik', fontsize=16,
fontweight='bold', y=1.00)

colors = ['skyblue', 'lightcoral', 'lightgreen', 'gold']

for idx, col in enumerate(numeric_cols):
    row = idx // 2
    col_pos = idx % 2

    # Histogram dengan KDE (Kernel Density Estimation)
    axes[row, col_pos].hist(df_clean[col], bins=30, color=colors[idx],
                            edgecolor='black', alpha=0.7, density=True)

    # Tambahkan KDE line
    df_clean[col].plot(kind='kde', ax=axes[row, col_pos], color='darkblue',
                       linewidth=2, secondary_y=False)

    axes[row, col_pos].set_title(f'Distribusi {col.upper()}', fontsize=12,
fontweight='bold')
    axes[row, col_pos].set_xlabel(col)
    axes[row, col_pos].set_ylabel('Density (Kepadatan)')
    axes[row, col_pos].grid(True, alpha=0.3)

    # Tambahkan garis mean dan median
    mean_val = df_clean[col].mean()
    median_val = df_clean[col].median()
    axes[row, col_pos].axvline(mean_val, color='red', linestyle='--',
linewidth=2, label=f'Mean: {mean_val:.2f}')
    axes[row, col_pos].axvline(median_val, color='green', linestyle='--',
linewidth=2,
                               label=f'Median: {median_val:.2f}')
    axes[row, col_pos].legend()

plt.tight_layout()
plt.savefig('histogram_distribusi.png', dpi=300, bbox_inches='tight')
plt.close()

print("✓ Histogram telah dibuat untuk semua variabel numerik")

# BOXPLOT untuk setiap variabel numerik
```

```
print("\n[B] BOXPLOT - Deteksi Outlier dan Persebaran Data")
print("-" * 80)

fig, axes = plt.subplots(2, 2, figsize=(14, 10))
fig.suptitle('BOXPLOT: Persebaran dan Outlier Variabel Numerik', fontsize=16,
fontweight='bold', y=1.00)

for idx, col in enumerate(numeric_cols):
    row = idx // 2
    col_pos = idx % 2

    # Boxplot dengan styling
    bp = axes[row, col_pos].boxplot(df_clean[col], vert=True,
patch_artist=True,
                                    boxprops=dict(facecolor=colors[idx],
alpha=0.7),
                                    medianprops=dict(color='red',
linewidth=2),
                                    whiskerprops=dict(color='black',
linewidth=1.5),
                                    capprops=dict(color='black',
linewidth=1.5),
                                    flierprops=dict(marker='o',
markerfacecolor='red',
                                                    markersize=6,
linestyle='none', alpha=0.5))

    axes[row, col_pos].set_title(f'Boxplot {col.upper()}', fontsize=12,
fontweight='bold')
    axes[row, col_pos].set_ylabel(col)
    axes[row, col_pos].grid(True, alpha=0.3, axis='y')

    # Tambahkan informasi statistik
    q1 = df_clean[col].quantile(0.25)
    q3 = df_clean[col].quantile(0.75)
    iqr = q3 - q1
    outliers = df_clean[(df_clean[col] < q1 - 1.5 * iqr) | (df_clean[col] > q3
+ 1.5 * iqr)][col]

    textstr = f'Q1: {q1:.2f}\nMedian: {df_clean[col].median():.2f}\nQ3:
{q3:.2f}\nOutliers: {len(outliers)}'
    axes[row, col_pos].text(1.15, 0.5, textstr, transform=axes[row,
col_pos].transAxes,
                            fontsize=10, verticalalignment='center',
                            bbox=dict(boxstyle='round', facecolor='wheat',
alpha=0.5))

plt.tight_layout()
```

```
plt.savefig('boxplot_distribusi.png', dpi=300, bbox_inches='tight')
plt.close()

print("✓ Boxplot telah dibuat untuk semua variabel numerik")

#
================================================================================
==
# KESIMPULAN STATISTIK DESKRIPTIF DAN UJI NORMALITAS
#
================================================================================
==

print("\n" + "=" * 80)
print("KESIMPULAN STATISTIK DESKRIPTIF DAN UJI NORMALITAS")
print("=" * 80)

print("""
KESIMPULAN:

1. STATISTIK DESKRIPTIF:
   a) Variabel Age (Umur):
      - Data tersebar cukup merata dari usia muda hingga tua
      - Rata-rata dan median relatif berdekatan

   b) Variabel BMI (Body Mass Index):
      - Mayoritas pemegang polis memiliki BMI antara 26-41
      - Distribusi cenderung simetris dengan sedikit skewness

   c) Variabel Children (Jumlah Anak):
      - Kebanyakan pemegang polis memiliki 0-2 anak
      - Distribusi right-skewed (miring ke kanan)

   d) Variabel Charges (Biaya Asuransi):
      - Sangat bervariasi dengan range yang lebar
      - Mean lebih besar dari median, menunjukkan distribusi right-skewed
      - Ada perbedaan signifikan antara biaya terendah dan tertinggi

2. UJI NORMALITAS (SHAPIRO-WILK):
   - Semua variabel numerik menunjukkan p-value < 0.05
   - Kesimpulan: SEMUA VARIABEL TIDAK MENGIKUTI DISTRIBUSI NORMAL
   - Ini umum terjadi pada data real-world, terutama untuk data biaya
     yang cenderung right-skewed

3. VISUALISASI:
   a) Histogram:
      - Menunjukkan pola distribusi setiap variabel
      - Charges menunjukkan distribusi multimodal (beberapa puncak)
```

```
      - Age relatif lebih uniform dibanding variabel lain

   b) Boxplot:
      - Terdeteksi beberapa outlier pada variabel charges
      - Outlier ini mungkin sah (orang dengan kondisi kesehatan khusus)
      - Variabel age dan bmi relatif tidak memiliki banyak outlier

4. IMPLIKASI:
   - Karena data tidak normal, penggunaan metode statistik non-parametrik
     mungkin lebih tepat untuk beberapa analisis
   - Namun, untuk regresi linear, asumsi normalitas residual lebih penting
     daripada normalitas variabel individual
   - Transformasi data (log, sqrt) bisa dipertimbangkan jika diperlukan
""")

#
==========================================================================
==
# 7. ANALISIS KORELASI
#
==========================================================================
==
# Menganalisis hubungan linear antara Age (umur) dengan Charges (biaya
asuransi)
# Korelasi mengukur kekuatan dan arah hubungan antara dua variabel

print("\n" + "=" * 80)
print("7. ANALISIS KORELASI")
print("=" * 80)

print("\n[A] KORELASI ANTARA AGE DAN CHARGES")
print("-" * 80)

# Menghitung korelasi Pearson
correlation_age_charges = df_clean['age'].corr(df_clean['charges'])
print(f"Koefisien Korelasi Pearson (r): {correlation_age_charges:.4f}")

# Interpretasi kekuatan korelasi
if abs(correlation_age_charges) < 0.3:
    strength = "LEMAH"
elif abs(correlation_age_charges) < 0.7:
    strength = "SEDANG"
else:
    strength = "KUAT"

direction = "POSITIF" if correlation_age_charges > 0 else "NEGATIF"

print(f"\nInterpretasi:")
```

```
print(f"  - Kekuatan: {strength}")
print(f"  - Arah: {direction}")
print(f"  - Artinya: Terdapat hubungan {direction.lower()} yang
{strength.lower()}")
print(f"            antara umur (age) dengan biaya asuransi (charges)")

# Matriks korelasi untuk semua variabel numerik
print("\n[B] MATRIKS KORELASI SEMUA VARIABEL NUMERIK")
print("-" * 80)

correlation_matrix = df_clean[numeric_cols].corr()
print(correlation_matrix)

# Visualisasi Heatmap Korelasi
print("\n[C] VISUALISASI HEATMAP KORELASI")
print("-" * 80)

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', center=0,
            square=True, linewidths=1, cbar_kws={"shrink": 0.8},
            fmt='.3f', vmin=-1, vmax=1)
plt.title('HEATMAP KORELASI VARIABEL NUMERIK', fontsize=14, fontweight='bold',
pad=20)
plt.tight_layout()
plt.savefig('heatmap_korelasi.png', dpi=300, bbox_inches='tight')
plt.close()

print("✓ Heatmap korelasi telah dibuat")

# Visualisasi Scatter Plot: Age vs Charges
print("\n[D] SCATTER PLOT: AGE vs CHARGES")
print("-" * 80)

plt.figure(figsize=(12, 6))
plt.scatter(df_clean['age'], df_clean['charges'], alpha=0.5, s=50,
c='steelblue', edgecolors='black', linewidth=0.5)
plt.xlabel('Age (Umur)', fontsize=12, fontweight='bold')
plt.ylabel('Charges (Biaya Asuransi)', fontsize=12, fontweight='bold')
plt.title(f'SCATTER PLOT: Hubungan Age vs Charges\n(Korelasi r =
{correlation_age_charges:.4f})',
          fontsize=14, fontweight='bold')
plt.grid(True, alpha=0.3)

# Tambahkan garis trend
z = np.polyfit(df_clean['age'], df_clean['charges'], 1)
p = np.poly1d(z)
plt.plot(df_clean['age'], p(df_clean['age']), "r--", linewidth=2,
label=f'Trend Line: y={z[0]:.2f}x+{z[1]:.2f}')
```

```
plt.legend()

plt.tight_layout()
plt.savefig('scatter_age_charges.png', dpi=300, bbox_inches='tight')
plt.close()

print("✓ Scatter plot telah dibuat")

#
================================================================================
==
# 8. ANALISIS REGRESI LINEAR
#
================================================================================
==
# Membuat model regresi linear untuk memprediksi Charges berdasarkan Age

print("\n" + "=" * 80)
print("8. ANALISIS REGRESI LINEAR")
print("=" * 80)

print("\n[A] MEMBANGUN MODEL REGRESI")
print("-" * 80)

# Membuat model regresi menggunakan statsmodels
model = smf.ols('charges ~ age', data=df_clean).fit()

# Ekstrak parameter model
intercept = model.params['Intercept']
slope = model.params['age']

print(f"Model Regresi Linear Sederhana:")
print(f"  Charges = beta0 + beta1 × Age")
print(f"\nParameter Model:")
print(f"  beta0 (Intercept/Konstanta): {intercept:,.2f}")
print(f"  beta1 (Slope/Koefisien Age): {slope:,.2f}")
print(f"\nPersamaan Regresi:")
print(f"  Charges = {intercept:,.2f} + {slope:,.2f} × Age")

print(f"\nInterpretasi:")
print(f"  - Intercept ({intercept:,.2f}): Biaya dasar asuransi ketika umur =
0")
print(f"  - Slope ({slope:,.2f}): Setiap kenaikan 1 tahun umur, biaya
asuransi")
print(f"                          meningkat sebesar ${slope:,.2f}")

# Ringkasan model
print("\n[B] RINGKASAN MODEL REGRESI")
```

```
print("-" * 80)
print(model.summary())

# Koefisien Determinasi (R-squared)
print("\n[C] KOEFISIEN DETERMINASI (R-SQUARED)")
print("-" * 80)

r_squared = model.rsquared
r_squared_adj = model.rsquared_adj

print(f"R-squared (R²): {r_squared:.4f}")
print(f"R-squared Adjusted: {r_squared_adj:.4f}")
print(f"\nInterpretasi R²:")
print(f"  - Model ini menjelaskan {r_squared * 100:.2f}% variasi dalam biaya
asuransi (charges)")
print(f"  - Sisanya ({(1 - r_squared) * 100:.2f}%) dijelaskan oleh faktor lain
yang tidak")
print(f"    termasuk dalam model (misalnya: status merokok, BMI, dll)")

if r_squared < 0.3:
    r2_interpretation = "RENDAH - Model kurang baik dalam menjelaskan variasi"
elif r_squared < 0.6:
    r2_interpretation = "SEDANG - Model cukup baik dalam menjelaskan variasi"
else:
    r2_interpretation = "TINGGI - Model sangat baik dalam menjelaskan variasi"

print(f"  - Tingkat kecocokan model: {r2_interpretation}")

# Sum of Squared Residuals (SSR)
print("\n[D] SUM OF SQUARED RESIDUALS (SSR)")
print("-" * 80)

ssr = model.ssr
mse = model.mse_resid
rmse = np.sqrt(mse)

print(f"SSR (Sum of Squared Residuals): {ssr:,.2f}")
print(f"MSE (Mean Squared Error): {mse:,.2f}")
print(f"RMSE (Root Mean Squared Error): {rmse:,.2f}")
print(f"\nInterpretasi SSR:")
print(f"  - SSR adalah total kuadrat kesalahan prediksi")
print(f"  - Semakin kecil SSR, semakin baik model dalam memprediksi")
print(f"  - RMSE ({rmse:,.2f}) menunjukkan rata-rata kesalahan prediksi")
print(f"    sekitar ${rmse:,.2f} dari nilai aktual")

# Visualisasi Model Regresi
print("\n[E] VISUALISASI MODEL REGRESI")
print("-" * 80)
```

```
fig, axes = plt.subplots(1, 2, figsize=(16, 6))

# Plot 1: Scatter plot dengan garis regresi
axes[0].scatter(df_clean['age'], df_clean['charges'], alpha=0.5, s=50,
                c='steelblue', edgecolors='black', linewidth=0.5, label='Data
Aktual')

# Garis regresi
x_pred = np.linspace(df_clean['age'].min(), df_clean['age'].max(), 100)
y_pred = intercept + slope * x_pred
axes[0].plot(x_pred, y_pred, 'r-', linewidth=3, label=f'Garis Regresi')

axes[0].set_xlabel('Age (Umur)', fontsize=12, fontweight='bold')
axes[0].set_ylabel('Charges (Biaya Asuransi)', fontsize=12, fontweight='bold')
axes[0].set_title(f'MODEL REGRESI LINEAR\nCharges = {intercept:,.2f} +
{slope:,.2f} × Age\nR² = {r_squared:.4f}',
                  fontsize=12, fontweight='bold')
axes[0].legend(loc='upper left', fontsize=10)
axes[0].grid(True, alpha=0.3)

# Plot 2: Residual plot
residuals = model.resid
axes[1].scatter(model.fittedvalues, residuals, alpha=0.5, s=50,
                c='coral', edgecolors='black', linewidth=0.5)
axes[1].axhline(y=0, color='red', linestyle='--', linewidth=2)
axes[1].set_xlabel('Fitted Values (Nilai Prediksi)', fontsize=12,
fontweight='bold')
axes[1].set_ylabel('Residuals (Kesalahan)', fontsize=12, fontweight='bold')
axes[1].set_title('RESIDUAL PLOT\n(Untuk Memeriksa Asumsi Model)',
fontsize=12, fontweight='bold')
axes[1].grid(True, alpha=0.3)

plt.tight_layout()
plt.savefig('model_regresi_linear.png', dpi=300, bbox_inches='tight')
plt.close()

print("✓ Visualisasi model regresi telah dibuat")

#
==========================================================================
==
# KESIMPULAN AKHIR ANALISIS KORELASI DAN REGRESI
#
==========================================================================
==

print("\n" + "=" * 80)
```

```
print("KESIMPULAN AKHIR ANALISIS KORELASI DAN REGRESI")
print("=" * 80)

print(f"""
RINGKASAN HASIL ANALISIS:

1. KORELASI ANTARA AGE DAN CHARGES:
   - Koefisien korelasi (r): {correlation_age_charges:.4f}
   - Kekuatan hubungan: {strength}
   - Arah hubungan: {direction}

   Interpretasi:
   Terdapat hubungan {direction.lower()} yang {strength.lower()} antara umur
(age)
   dengan biaya asuransi (charges). Ini berarti:
   → Semakin tua umur seseorang, biaya asuransi cenderung meningkat
   → Namun, hubungan ini tidak terlalu kuat karena ada faktor lain yang
     juga mempengaruhi biaya asuransi (seperti status merokok, BMI, dll)

2. MODEL REGRESI LINEAR:
   Persamaan: Charges = {intercept:,.2f} + {slope:,.2f} × Age

   Interpretasi:
   a) Intercept ({intercept:,.2f}):
      - Biaya dasar asuransi (ketika umur = 0)
      - Nilai ini adalah estimasi teoritis

   b) Slope ({slope:,.2f}):
      - Setiap bertambah 1 tahun umur, biaya asuransi naik ${slope:,.2f}
      - Contoh: Umur 30 tahun vs 40 tahun
        * Umur 30: ${intercept + slope * 30:,.2f}
        * Umur 40: ${intercept + slope * 40:,.2f}
        * Selisih: ${slope * 10:,.2f}

3. KOEFISIEN DETERMINASI (R²):
   - R² = {r_squared:.4f} atau {r_squared * 100:.2f}%
   - R² Adjusted = {r_squared_adj:.4f}

   Interpretasi:
   → Model ini hanya menjelaskan {r_squared * 100:.2f}% variasi dalam biaya
asuransi
   → Sisanya ({(1 - r_squared) * 100:.2f}%) dipengaruhi oleh faktor lain
seperti:
     * Status merokok (smoker)
     * Body Mass Index (BMI)
     * Jumlah anak (children)
     * Wilayah tempat tinggal (region)
     * Riwayat penyakit
```

```
     * dll
   → Nilai R² yang {strength.lower()} menunjukkan bahwa model sederhana ini
     belum optimal untuk prediksi

4. SUM OF SQUARED RESIDUALS (SSR):
   - SSR: {ssr:,.2f}
   - RMSE: {rmse:,.2f}

   Interpretasi:
   → Rata-rata kesalahan prediksi model adalah sekitar ${rmse:,.2f}
   → Ini cukup besar mengingat range biaya asuransi yang lebar
   → Menunjukkan bahwa prediksi hanya berdasarkan umur saja tidak cukup akurat

5. REKOMENDASI:
   a) Model Sederhana vs Model Kompleks:
      - Model regresi sederhana (hanya age) kurang optimal
      - Disarankan menggunakan multiple regression dengan variabel tambahan:
        * Age + BMI + Smoker + Children + Region
      - Model yang lebih kompleks akan meningkatkan R² dan akurasi prediksi

   b) Aplikasi Praktis:
      - Umur adalah faktor penting tapi bukan satu-satunya penentu
      - Perusahaan asuransi harus mempertimbangkan banyak faktor
      - Status merokok kemungkinan memiliki pengaruh lebih besar daripada umur

   c) Validasi Model:
      - Residual plot menunjukkan pola yang perlu dievaluasi
      - Perlu dilakukan uji asumsi regresi lebih lanjut:
        * Linearitas
        * Homoskedastisitas
        * Normalitas residual
        * Independensi residual

KESIMPULAN UTAMA:
================================================================================
==
Meskipun terdapat hubungan positif antara umur (age) dengan biaya asuransi
(charges), hubungan ini tergolong {strength.lower()} dan hanya menjelaskan
{r_squared * 100:.2f}%
variasi dalam biaya asuransi. Oleh karena itu, untuk prediksi yang lebih
akurat,
diperlukan model yang lebih kompleks dengan memasukkan variabel-variabel lain
yang relevan seperti status merokok, BMI, dan faktor kesehatan lainnya.

Model sederhana ini dapat digunakan sebagai baseline atau untuk memberikan
gambaran umum tentang pengaruh umur terhadap biaya asuransi, namun tidak
disarankan untuk digunakan sebagai satu-satunya dasar dalam pengambilan
keputusan bisnis yang kritis.
```

```
================================================================================
==
""")

print("\n" + "=" * 80)
print("ANALISIS SELESAI")
print("=" * 80)
print("\n✓ Semua analisis telah selesai dilakukan")
print("✓ File data bersih: insurance_remed_clean.xlsx")
print("✓ File visualisasi:")
print("  - boxplot_outlier_detection.png")
print("  - histogram_distribusi.png")
print("  - boxplot_distribusi.png")
print("  - heatmap_korelasi.png")
print("  - scatter_age_charges.png")
print("  - model_regresi_linear.png")
print("=" * 80)
```