

Initial Data Analysis Report

1. Dataset Info

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 200 entries, 0 to 199

Data columns (total 75 columns):

#	Column	Non-Null Count	Dtype
---	--------	----------------	-------

--- -----

0	id2	200 non-null	int64
1	universi	199 non-null	object
2	region	199 non-null	object
3	cluster	199 non-null	float64
4	cross	199 non-null	object
5	sex	199 non-null	object
6	birth_ye	199 non-null	float64
7	sample_c	199 non-null	float64
8	h_18	178 non-null	float64
9	h_19	184 non-null	float64
10	h_20	192 non-null	object
11	h_21	187 non-null	object
12	h_22	179 non-null	float64
13	h_23	187 non-null	float64
14	h_24	191 non-null	object
15	h_25	190 non-null	object
16	h_26	193 non-null	float64
17	h_27	189 non-null	float64
18	h_28	193 non-null	object
19	h_29	193 non-null	object
20	h_30	199 non-null	object
21	h_31	198 non-null	object
22	h_32	197 non-null	object
23	h_33	193 non-null	object
24	h_34	196 non-null	object
25	h_35_a	192 non-null	object
26	h_35_b	104 non-null	object
27	h_36	192 non-null	object
28	h_37	197 non-null	object
29	h_38	196 non-null	object
30	h_39	198 non-null	object
31	h_40	196 non-null	object
32	h_41	198 non-null	object
33	h_42	198 non-null	object
34	h_43_1	195 non-null	object
35	h_43_2	196 non-null	object
36	h_44_1	195 non-null	object
37	h_44_2	197 non-null	object
38	h_44_3	195 non-null	object
39	h_45	197 non-null	object
40	z_52	191 non-null	object
41	z_53	198 non-null	object
42	z_54	194 non-null	object

43 z_55 195 non-null object
44 z_56_1 187 non-null object
45 z_56_2 186 non-null object
46 z_56_3 186 non-null object
47 z_56_4 181 non-null object
48 z_56_5 189 non-null object
49 z_56_6 191 non-null object
50 z_56_7 184 non-null object
51 h_57 198 non-null object
52 h_58 197 non-null object
53 h_59 194 non-null object
54 h_60 191 non-null object
55 h_61 191 non-null object
56 h_62 186 non-null object
57 h_63_1 185 non-null float64
58 h_63_2 184 non-null float64
59 h_64_1 186 non-null float64
60 h_64_2 188 non-null float64
61 sumScore 163 non-null float64
62 tertiles 163 non-null object
63 n_82 196 non-null object
64 n_83 191 non-null object
65 n_84 192 non-null object
66 n_85 191 non-null object
67 n_86 197 non-null object
68 n_87 193 non-null object
69 n_88 189 non-null object
70 n_89 192 non-null object
71 n_90 193 non-null object
72 n_91 197 non-null object
73 n_92 197 non-null object
74 n_93 192 non-null float64

dtypes: float64(15), int64(1), object(59)

memory usage: 117.3+ KB

2. Descriptive Statistics

Feature	count	mean	std	min	25%	50%	75%	max
id2	200.0	1111944304.96	6727539.94	1101118501.0	1106118306.75	1112218004.5	1118217801.5	1123128702.0
cluster	199.0	11.78	6.73	1.0	6.0	12.0	18.0	23.0
birth_ye	199.0	82.14	2.81	76.0	80.0	82.0	84.0	87.0
sample_c	199.0	5.48	2.88	1.0	3.0	6.0	8.0	10.0
h_18	178.0	7.8	1.16	6.0	7.0	7.3	9.0	12.0
h_19	184.0	9.23	0.97	7.0	9.0	9.0	10.0	12.0
h_22	179.0	13.94	0.91	11.0	13.3	14.0	14.3	16.0
h_23	187.0	14.05	0.7	12.0	14.0	14.0	14.3	16.0
h_26	193.0	20.9	0.92	19.0	20.3	21.0	21.3	24.0
h_27	189.0	21.01	0.96	19.0	20.3	21.0	21.3	24.0
h_63_1	185.0	8.3	1.35	4.0	8.0	8.0	9.0	15.0

h_63_2	184.0	22.61	1.08	19.0	22.0	23.0	23.0	25.3
h_64_1	186.0	9.2	1.52	4.0	8.0	9.0	10.0	14.0
h_64_2	188.0	23.12	1.12	19.0	22.3	23.0	24.0	26.0
sumScore	163.0	20.53	8.78	11.0	14.0	18.0	26.0	49.0
n_93	192.0	8.48	2.07	0.0	8.0	9.0	10.0	10.0

3. Missing Values per Column

id2: 0
universi: 1
region: 1
cluster: 1
cross: 1
sex: 1
birth_ye: 1
sample_c: 1
h_18: 22
h_19: 16
h_20: 8
h_21: 13
h_22: 21
h_23: 13
h_24: 9
h_25: 10
h_26: 7
h_27: 11
h_28: 7
h_29: 7
h_30: 1
h_31: 2
h_32: 3
h_33: 7
h_34: 4
h_35_a: 8
h_35_b: 96
h_36: 8
h_37: 3
h_38: 4
h_39: 2
h_40: 4
h_41: 2
h_42: 2
h_43_1: 5
h_43_2: 4

h_44_1: 5
h_44_2: 3
h_44_3: 5
h_45: 3
z_52: 9
z_53: 2
z_54: 6
z_55: 5
z_56_1: 13
z_56_2: 14
z_56_3: 14
z_56_4: 19
z_56_5: 11
z_56_6: 9
z_56_7: 16
h_57: 2
h_58: 3
h_59: 6
h_60: 9
h_61: 9
h_62: 14
h_63_1: 15
h_63_2: 16
h_64_1: 14
h_64_2: 12
sumScore: 37
tertiles: 37
n_82: 4
n_83: 9
n_84: 8
n_85: 9
n_86: 3
n_87: 7
n_88: 11
n_89: 8
n_90: 7
n_91: 3
n_92: 3
n_93: 8