

# DATA QUALITY REPORT

## Null Checks & Data Validation

- There are no null values, empty strings, white spaces, and blank values in the data.
- As we can see in the below table that there are outliers and extremes in the data.

	Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	Complete %	Valid Records	Null Value	Empty String	White Space	Blank Value
1	Q1A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
2	Q1I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
3	Q1E	Continuous	14	21	None	Never	Fixed	100	39775	0	0	0	0
4	Q2A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
5	Q2I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
6	Q2E	Continuous	32	32	None	Never	Fixed	100	39775	0	0	0	0
7	Q3A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
8	Q3I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
9	Q3E	Continuous	9	13	None	Never	Fixed	100	39775	0	0	0	0
10	Q4A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
11	Q4I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
12	Q4E	Continuous	13	26	None	Never	Fixed	100	39775	0	0	0	0
13	Q5A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
14	Q5I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
15	Q5E	Continuous	13	24	None	Never	Fixed	100	39775	0	0	0	0
16	Q6A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
17	Q6I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
18	Q6E	Continuous	21	34	None	Never	Fixed	100	39775	0	0	0	0
19	Q7A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
20	Q7I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
21	Q7E	Continuous	1	6	None	Never	Fixed	100	39775	0	0	0	0
22	Q8A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
23	Q8I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
24	Q8E	Continuous	32	35	None	Never	Fixed	100	39775	0	0	0	0
25	Q9A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
26	Q9I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
27	Q9E	Continuous	2	5	None	Never	Fixed	100	39775	0	0	0	0
28	Q10A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
29	Q10I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
30	Q10E	Continuous	4	10	None	Never	Fixed	100	39775	0	0	0	0
31	Q11A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0

PROJECT – I DEPRESSION ANXIETY STRESS SCALES PREDICTION	KRUPAL PATEL (20162121007) YASH PRAJAPATI (20162121023)
---	--

32	Q11I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
33	Q11E	Continuous	11	25	None	Never	Fixed	100	39775	0	0	0	0
34	Q12A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
35	Q12I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
36	Q12E	Continuous	6	14	None	Never	Fixed	100	39775	0	0	0	0
37	Q13A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
38	Q13I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
39	Q13E	Continuous	9	25	None	Never	Fixed	100	39775	0	0	0	0
40	Q14A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
41	Q14I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
42	Q14E	Continuous	4	7	None	Never	Fixed	100	39775	0	0	0	0
43	Q15A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
44	Q15I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
45	Q15E	Continuous	17	26	None	Never	Fixed	100	39775	0	0	0	0
46	Q16A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
47	Q16I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
48	Q16E	Continuous	9	22	None	Never	Fixed	100	39775	0	0	0	0
49	Q17A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
50	Q17I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
51	Q17E	Continuous	19	26	None	Never	Fixed	100	39775	0	0	0	0
52	Q18A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
53	Q18I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
54	Q18E	Continuous	2	9	None	Never	Fixed	100	39775	0	0	0	0
55	Q19A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
56	Q19I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
57	Q19E	Continuous	8	11	None	Never	Fixed	100	39775	0	0	0	0
58	Q20A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
59	Q20I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
60	Q20E	Continuous	25	24	None	Never	Fixed	100	39775	0	0	0	0
61	Q21A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
62	Q21I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
63	Q21E	Continuous	3	8	None	Never	Fixed	100	39775	0	0	0	0
64	Q22A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
65	Q22I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
66	Q22E	Continuous	52	46	None	Never	Fixed	100	39775	0	0	0	0
67	Q23A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
68	Q23I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
69	Q23E	Continuous	94	73	None	Never	Fixed	100	39775	0	0	0	0
70	Q24A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0

PROJECT – I DEPRESSION ANXIETY STRESS SCALES PREDICTION	KRUPAL PATEL (20162121007) YASH PRAJAPATI (20162121023)
---	--

71	Q24I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
72	Q24E	Continuous	5	8	None	Never	Fixed	100	39775	0	0	0	0
73	Q25A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
74	Q25I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
75	Q25E	Continuous	11	18	None	Never	Fixed	100	39775	0	0	0	0
76	Q26A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
77	Q26I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
78	Q26E	Continuous	3	22	None	Never	Fixed	100	39775	0	0	0	0
79	Q27A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
80	Q27I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
81	Q27E	Continuous	3	9	None	Never	Fixed	100	39775	0	0	0	0
82	Q28A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
83	Q28I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
84	Q28E	Continuous	1	2	None	Never	Fixed	100	39775	0	0	0	0
85	Q29A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
86	Q29I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
87	Q29E	Continuous	4	13	None	Never	Fixed	100	39775	0	0	0	0
88	Q30A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
89	Q30I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
90	Q30E	Continuous	12	12	None	Never	Fixed	100	39775	0	0	0	0
91	Q31A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
92	Q31I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
93	Q31E	Continuous	21	29	None	Never	Fixed	100	39775	0	0	0	0
94	Q32A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
95	Q32I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
96	Q32E	Continuous	8	12	None	Never	Fixed	100	39775	0	0	0	0
97	Q33A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
98	Q33I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
99	Q33E	Continuous	10	27	None	Never	Fixed	100	39775	0	0	0	0
100	Q34A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
101	Q34I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
102	Q34E	Continuous	4	12	None	Never	Fixed	100	39775	0	0	0	0
103	Q35A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
104	Q35I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
105	Q35E	Continuous	4	15	None	Never	Fixed	100	39775	0	0	0	0
106	Q36A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
107	Q36I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
108	Q36E	Continuous	22	32	None	Never	Fixed	100	39775	0	0	0	0
109	Q37A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0

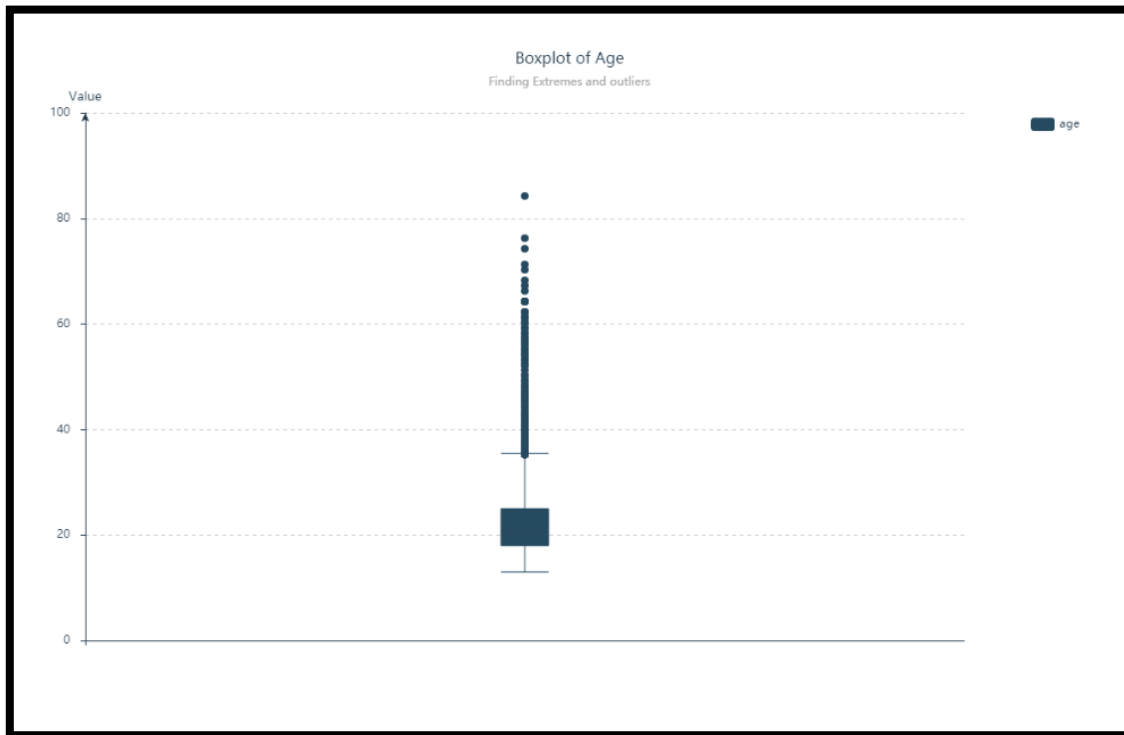
110	Q37I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
111	Q37E	Continuous	10	22	None	Never	Fixed	100	39775	0	0	0	0
112	Q38A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
113	Q38I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
114	Q38E	Continuous	2	2	None	Never	Fixed	100	39775	0	0	0	0
115	Q39A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
116	Q39I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
117	Q39E	Continuous	9	14	None	Never	Fixed	100	39775	0	0	0	0
118	Q40A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
119	Q40I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
120	Q40E	Continuous	4	14	None	Never	Fixed	100	39775	0	0	0	0
121	Q41A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
122	Q41I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
123	Q41E	Continuous	12	16	None	Never	Fixed	100	39775	0	0	0	0
124	Q42A	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
125	Q42I	Continuous	0	0	None	Never	Fixed	100	39775	0	0	0	0
126	Q42E	Continuous	20	33	None	Never	Fixed	100	39775	0	0	0	0
127	countr y	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
128	sourc e	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
129	introel apse	Continuous	4	17	None	Never	Fixed	100	39775	0	0	0	0
130	testel apse	Continuous	0	18	None	Never	Fixed	100	39775	0	0	0	0
131	surve yelaps e	Continuous	15	29	None	Never	Fixed	100	39775	0	0	0	0
132	TIPI1	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
133	TIPI2	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
134	TIPI3	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
135	TIPI4	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
136	TIPI5	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
137	TIPI6	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
138	TIPI7	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
139	TIPI8	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
140	TIPI9	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
141	TIPI1 0	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
142	VCL1	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
143	VCL2	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0

PROJECT – 1 DEPRESSION ANXIETY STRESS SCALES PREDICTION	KRUPAL PATEL (20162121007) YASH PRAJAPATI (20162121023)
---	--

144	VCL3	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
145	VCL4	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
146	VCL5	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
147	VCL6	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
148	VCL7	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
149	VCL8	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
150	VCL9	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
151	VCL10	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
152	VCL11	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
153	VCL12	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
154	VCL13	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
155	VCL14	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
156	VCL15	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
157	VCL16	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
158	education	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
159	urban	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
160	gender	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
161	engnat	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
162	age	Continuous	4	5	None	Never	Fixed	100	39775	0	0	0	0
163	screen size	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
164	uniquenetwork location	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
165	hand	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
166	religion	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
167	orientation	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
168	race	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
169	voted	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
170	married	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0
171	family size	Nominal	--	--	--	Never	Fixed	100	39775	0	0	0	0

## Outliers & Extremes

- **Age:** There are several outliers and extremes present in the data as shown in the boxplot below.



## Spelling Inconsistencies

- There are spelling inconsistency in the field of 'major'.

## Plausibility Check

- There is plausibility in the data. For example, the age of a living person cannot be 1998.

## Non-Priority Fields

There are some fields that do not impact the modelling based on our hypothesis. They are:

- **QxE, introelapse, testelapse, surveyelapse** – Time taken to answer the question.
- **QxI** – The order of the questions.
- **engnat** – Native language of any person.
- **hand** – Strong hand of any person.

- **orientation** – Sexual orientation of a person.
- **voted** – Voting status of person.
- **screen size** – Screen size of device.
- **unique network location** – Network location of the user.
- **source** – How user found the test.
- **VCLx** - The words whose definitions the person is sure he/she knows.
- **country**

### Data Inconsistency

Our data is stored in csv, so there is no inconsistency among the records. All the fields have same number of records.