EDA

Last Update: March 2, 2015

1 Summary Statistics

1.1 General Information

LIST_OWN and LIST_BACK are very left-skewed, without any sort of a bell shape to them. AGE is also left-skewed, but does have a bell shape to it. Gender is roughly evenly split

	RESPID	GENDER	AGE	LIST_OWN	LIST_BACK
Min.	0	0.0000	13.00	0.000	0.0
1st Qu.	12161	0.0000	25.00	0.000	1.0
Median	26605	0.0000	39.00	1.000	2.0
Mean	25562	0.4963	39.28	2.177	2.9
3rd Qu.	38766	1.0000	52.00	3.000	4.0
Max.	50927	1.0000	94.00	16.000	16.0
NA's				10	10

1.2 Questions

All of the histograms for the questions have a periodicity to them with clear spikes at intervals of 25. 1-3 and 9-14 look the most normally distributed, 4-8, 15, 16, 18, and 19 are more to the left and 17 is a little to the right.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1st Qu.	28.00	44.00	31.00	14.00	12.00	14.00	10.00	9.00	28.00	41.00
Median	51.00	53.00	52.00	34.00	32.00	35.00	30.00	23.00	50.00	53.00
Mean	49.11	54.62	51.28	37.31	34.59	39.33	33.85	29.16	47.83	55.01
3rd Qu.	70.00	71.00	71.00	52.00	51.00	53.00	52.00	49.00	70.00	72.00
Max.	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NA's										

	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1st Qu.	48.00	36.00	28.00	33.00	13.00	11.00	35.00	17.00	14.00
Median	64.00	54.00	50.00	53.00	37.00	32.00	56.00	47.00	45.00
Mean	58.64	53.67	46.96	53.45	39.66	35.58	53.83	42.23	41.36
3rd Qu.	73.00	71.00	68.00	71.00	56.00	52.00	71.00	58.00	57.00
Max.	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
NA's						6435		13125	13125

1.3 Working

Key	Value
Working 0	WORKING.
Working 1	WORKING.Employed 30+ hours a week
Working 2	WORKING.Employed 8-29 hours per week
Working 3	WORKING.Employed part-time less than 8 hours per week
Working 4	WORKING.Full-time housewife / househusband
Working 5	WORKING.Full-time student
Working 6	WORKING.In unpaid employment (e.g. voluntary work)
Working 7	WORKING.Other
Working 8	WORKING.Part-time student
Working 9	WORKING.Prefer not to state
Working 10	WORKING.Retired from full-time employment (30+ hours per week)
Working 11	WORKING.Retired from self-employment
Working 12	WORKING.Self-employed
Working 13	WORKING.Temporarily unemployed

	Working 0	Working 1	Working 2	Working 3	Working 4	Working 5	Working 6
Min.	0.0000	0.0000	0.000	0.000000	0.000	0.0000	0.00000
1st Qu	0.0000	0.0000	0.000	0.000000	0.000	0.0000	0.00000
Median	0.0000	0.0000	0.000	0.000000	0.000	0.0000	0.00000
Mean	0.2698	0.2799	0.084	0.007914	0.054	0.1049	0.00701
3rd Qu.	1.0000	1.0000	0.000	0.000000	0.000	0.0000	0.00000
Max.	1.0000	1.0000	1.000	1.000000	1.000	1.0000	1.00000
NA's							

	Working 7	Working 8	Working 9	Working 10	Working 11	Working 12	Working 13
Min.	0.00000	0.000000	0.000000	0.00000	0.000000	0.00000	0.00000
1st Qu.	0.00000	0.000000	0.000000	0.00000	0.000000	0.00000	0.00000
Median	0.00000	0.000000	0.000000	0.00000	0.000000	0.00000	0.00000
Mean	0.02905	0.004502	0.004831	0.06767	0.007894	0.03965	0.03879
3rd Qu.	0.00000	0.000000	0.000000	0.00000	0.000000	0.00000	0.00000
Max.	1.00000	1.000000	1.000000	1.00000	1.000000	1.00000	1.00000
NA's							

1.4 Region

Key	Value
Region 0	REGION.
Region 1	REGION.Centre
Region 2	REGION.Midlands
Region 3	REGION.North
Region 4	REGION.North Ireland
Region 5	REGION.Northern Ireland
Region 6	REGION.South

	Region 0	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
Min.	0.00000	0.00000	0.0000	0.000	0.000000	0.00000	0.0000
1st Qu.	0.00000	0.00000	0.0000	0.0000	0.000000	0.00000	0.0000
Median	0.00000	0.00000	0.0000	0.0000	0.000000	0.00000	0.0000
Mean	0.02138	0.05851	0.2435	0.3441	0.002837	0.01581	0.3138
3rd Qu.	0.00000	0.00000	0.0000	1.0000	0.000000	0.00000	1.0000
Max.	1.00000	1.00000	1.0000	1.0000	1.000000	1.00000	1.0000
NA's							

1.5 Music

Key	Value
Music 1	MUSIC.I like music but it does not feature heavily in my life
Music 2	MUSIC.Music has no particular interest for me
Music 3	MUSIC.Music is important to me but not necessarily more important
Music 4	MUSIC.Music is important to me but not necessarily more important than other hobbies or interests
Music 5	MUSIC.Music is no longer as important as it used to be to me
Music 6	MUSIC.Music means a lot to me and is a passion of mine

	Music 1	Music 2	Music 3	Music 4	Music 5	Music 6
Min.	0.0000	0.00000	0.0000	0.00000	0.00000	0.00
1st Qu.	0.0000	0.00000	0.0000	0.0000	0.00000	0.00
Median	0.0000	0.00000	0.0000	0.00000	0.0000	0.00
Mean	0.2424	0.02132	0.3425	0.05084	0.03297	0.31
3rd Qu.	0.0000	0.00000	1.0000	0.00000	0.00000	1.00
Max.	1.0000	1.00000	1.00000	1.00000	1.00000	1.00
NA's						

2 Training and Testing Data in Users

Number of Training Rows Not In Users	11,857
Number of Testing Rows Not In Users	7,867
Proportion of Training Not In Users	0.06283952
Proportion of Testing Not In Users	0.06253875
Number of Unique Training Users Not In Users	2278
Number of Unique Testing Users Not In Users	2227
Proportion of Unique Training Not In Users	0.04603973
Proportion of Unique Testing Not In Users	0.04831641

3 Words File

The words.csv file contains 85 adjectives which may or may not describe a given artist. Of these, only 25 do not have NA for over 10% of the responses.

HEARD_OF	OWN_ARTIST_MUSIC			
Youthful	Current	Edgy	Thoughtful	
Fun	Authentic	Stylish	Beautiful	
Catchy	Passionate	Credible	Cool	
Talented	Distinctive	Timeless	Original	
Upbeat	Energetic	Approachable	Trendsetter	
Warm		None.of.these	Sexy	

Table 1: Attributes of words.csv file which are at least 90% real values (fewer than 11,830 rows are NA). Bold items contain NA values.

4 Histograms