

Total No of Questions (of each category): 10321

Occurrences of a few phrases in various categories:-

Word	Original Question	Logeq	Necc	Mutex
<i>Anything</i>	0	1	793	50
<i>Anyone</i>	0	0	64	0
<i>Is there</i>	31	875	3635	679
Is it	105	99	140	99
Is this	881	865	1067	861
<i>There a</i>	16	404	4172	442
one	105	196	162	213
<i>1</i>	0	440	5	11
no	103	325	141	359
color	1822	186	135	177
<i>0</i>	0	104	1	0
Are there	376	1084	886	1062
<i>Are these</i>	151	149	268	150
Are they	138	153	143	153
number	125	95	85	94
made	164	167	27	169
<i>animal</i>	396	225	405	233
2	0	504	4	105
3	0	213	2	231
4	0	150	2	289
5	1	82	1	307
6	1	79	1	195

7	1	36	1	133
8	0	34	1	147
9	0	12	0	71
red	83	321	77	357
blue	40	268	41	502
green	59	227	48	257
yellow	43	163	43	182
orange	30	110	29	92
black	52	313	46	371
white	99	525	84	305

We can see that the questions which contain the words written in *red* may have a high bias towards a particular answer.

Differences observed between our dataset and Facebook and ConVQA datasets and other possible biases:-

- 1) We have observed that our dataset does not have any complex questions. They mostly consist of changes in word orders, addition of conjunctions, incorporation of the answer into the questions and replacing a few words with similar(or opposite) words.
- 2) Unlike the other datasets, our dataset does not incorporate common sense in the questions. For example, the other datasets convert the question "Is this red?" into a question "Is this the same color as a tomato?". Our dataset does not have such questions.
- 3) Our questions only have binary yes and no answers, unlike the other two datasets which can have any word as an answer.
- 4) We have also observed that Necc category questions, which have "yes" as an answer are on an average shorter than the other category questions. They also seem to have lesser information in general.
- 5) Other datasets have rephrasings of the same question which have the same implication and the same answer. For example, a question "Where is the nike sign?" can have a rephrasing "Where is the nike sign located?". Our dataset has binary answer questions for each question.

- 6) Our dataset does not incorporate an attention mechanism and relies only on the question to form consistent questions.