Error Analysis: Select Column Test Set

Column-wise Statistics

	True 1	True 0
Pred 1	54	107
Pred 0	12	386

Column-wise Statistics

Metric	Value
Accuracy	78.71
Precision	33.54
Recall	81.81

Query-wise Statistics

Metric	Value
Accuracy	9.38

Questions Tagged Incorrectly

ID	Question	Marked As	Should Be?
279	What is the treatment for Epstein Barr virus	entity	desc
289	How much are egg mcmuffins	count	money
300	What years are millennials considered born in	desc	date

Query-wise Statistics

Question Type	Correct	Total	Wrong IDs
Desc	2: 250,279	13	248,249,251,255,258,259,264,271,282,283,284,300
Entity	0	7	240,242,254,268,270,276,288
Loc	2: 252,274	8	257,265,272,280,286,301
Date	0	19	238,239,243,246,247,260,262,263,266,269,275,277,281,290, 291,293,296,297,300
Count	1: 244	7	241,253,278,295,298,299
Money	0	2	256,289
Other Num	0	3	245,292,294
Yes/No	1:261	5	267,273,,285,287

Error Analysis

- Overall, same type of problems as last week, just a little less intense
- Over-predicting 1s, especially if there are multiple columns of the same type in a table
- For question types with few examples like money and other number, the model doesn't seem to have learnt anything
- The model seems to be confusing Yes/No questions with Desc(text) type