Question 3:

3.1 Analytics:

Per my testing in the current assignment, I think CRF (Conditional Random Fields) model works better than logistic regression for our sequence tagging problem.

Logistic Regression: When using a logistic regression model, we assume complete independence between the different components yi in y. The classifier, when predicting the label yi, can use any features of the input x and the position i. It is sometimes more intuitive to think about the features as being computed for the observed input, but then picking the corresponding parameters per the yi label you are trying to predict.

Conditional Random fields: Conditional random fields are an extension of logistic regression that incorporates sequential information in the labels, while still supporting the use of arbitrary features. The score function for a conditional random field thus combines both the evidence from the observed tokens and from the neighboring tags.
Predicting the best sequence from a CRF is not as straightforward as in logistic regression. We need a step such as Viterbi Algorithm to find the best sequence.

As per me, crf model works better in sequence tagging problem, mainly because this model will account for the lables of the previous observed inputs as well the features of the current input using Viterbi algortihm. On the other hand, the logistic regression, will only take features of each token into account at a time, making it more susceptible to missing out on context of the word.
Taking an example, in the NER problem, we find a word, “Chicago Bulls in the top of the league – the logistic regression will take Chicago as a place and name it as B-geo-loc, on the other hand, crf would consider the context by taking the entire sequence and it would appropriately tag Chicago Bulls as B-sportsteam, I-sportsteam respectively.