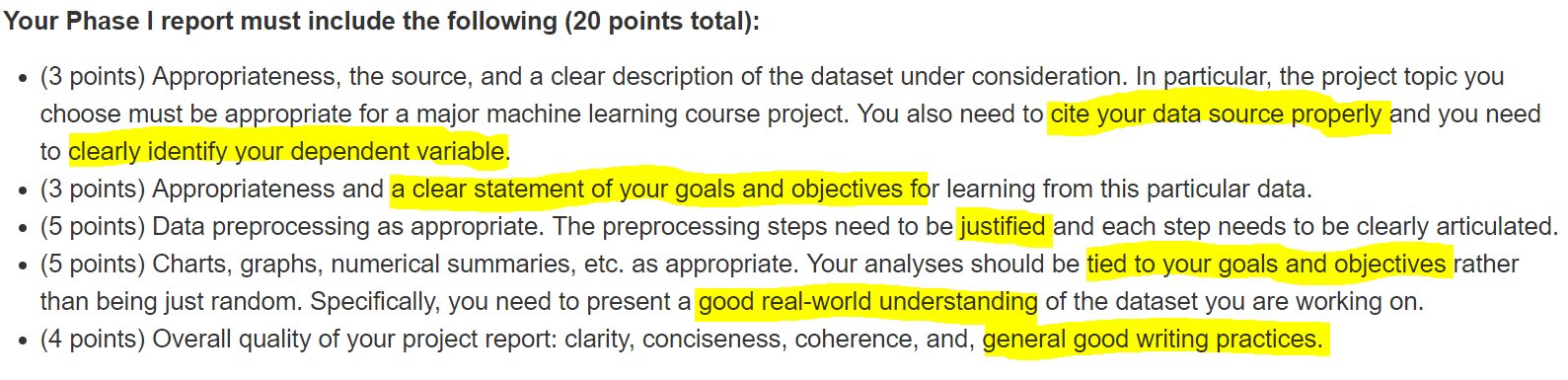
Ideas for ML Assignment Phase 1

* Vural mentioned in the lecture that the data is already pretty clean (no missing values or outliers) – can test anyway to confirm?
* A lot of the features have already been normalised between 0 and 1 but price could potentially still be normalised
* Time data: day 1 to 30 and also day of the week – does this need special treatment?
* Some of the metrics I don’t quite understand, especially those ratios. And what is meant by the target variable y is a numeric revenue-related metric?
* Map out relationships between variables
* Chapter 2 of the text book is all about: what is the business problem that we are trying to solve with machine learning?
  + Want to increase sales revenue from advertising by better targeting users with ads for things they are more likely to buy?
* Chapter 3:
  + Check distribution and cardinality(?) of data
  + visualise relationship between features: box plots, histogram
  + Covariance and correlation between features, scatter plot matrix
  + Binning

Marking Rubric



Key Things From Course Project Description

