

1. A marketing firm plans to pay people to make positive tweets about a client's product. The firm has data about the originator and content of previous tweets, the number of times each was retweeted, and the total number of recipients of a tweet (initially plus retweets). The firm does not have reliable past data about followers; although they can know current followers, they do not know who each person's followers were back at the time of each past tweet.
 - (a) For each of the questions below, select a model/approach that the could firm use. **[NOTE: CHOICES WILL BE THINGS LIKE "Exponential smoothing", "Logistic regression", etc.]**
 - i. [3 points] What model/approach could the firm use to predict the total number of recipients of a new tweet from the data based on content and originator?
 - ii. [3 points] What model/approach could the firm use to reduce the number of predictors used?
 - iii. [3 points] Suppose the firm was concerned that the diluting effect of political tweets made more-recent data different from older data. What model/approach could the firm use to determine whether, and when (if at all), the change happened?
 - iv. [3 points] Rather than analyze each product and each tweeter individually, the firm would like to group products and people. What model/approach could the firm use to determine how to make the groupings?
 - v. [3 points] The firm would like to initially pay a larger number of tweeters, and the, after observing their success, gradually reduce the number of people being paid. What model/approach could the firm use to go from a larger to a smaller number of tweeters over time?
 - (b) [2 points each, 10 total] Redo part (a) above, by giving a different answer for each part (i)-(v). [There is more than one good answer for each.]
 - (c) [3 points] It is likely that multiple tweeters' tweets could eventually reach the same recipient. What model/approach could the firm use to select tweeters in a way that maximizes the number of people reached by at least one of them?

2. Every morning at 10:30am, the manager of a fast-food restaurant determines how many hamburgers to pre-make so they are ready to be immediately given to customers at lunchtime. If not enough are pre-made, customers will have to wait a long time in line, and might go to the competing fast-food restaurant next door instead. If too many are pre-made, some will spoil before they can be used.

The manager has come up with two ideas:

- i. **Given** the past fraction of days π_k that k hamburgers were ordered at lunch, the past overall distributions of (a) arrival rates of customers, (b) time it takes to give a customer a pre-made hamburger, (c) time it takes to make and give the customer a newly-made hamburger, (d) customers leaving because of lines, and (e) hamburger spoilage, **use** a simulation model **to** find the best tradeoff between the expected number of customers served each day and the cost of spoiled hamburgers.
 - ii. **Given** past data on the probability p_{kn} that n hamburgers will be ordered at lunch today given that k were ordered yesterday, **use** a Markov chain model **to** find the steady-state probabilities π_k that there are k hamburgers ordered in a day. Then, **given** these probabilities and the relative costs of losing customers or having hamburgers spoil, **use** an optimization model **to** determine the most cost-effective number of hamburgers to order each day.
3. [6 points] Briefly select one problem with each approach (i)-(ii). **[NOTE: Choices will be provided.]**
4. [15 points] Select data/models/questions that you would combine to approach this question. The choices will be provided in this format:

Given _____{DATA}_____, use _____{MODEL(S)}_____ to _____{QUESTION/ANSWER}_____.

