Intellect Design Arena Data Scientist Recruitment Tasks

1. Two stock analysts writes daily reports on Apple stock. Their report consists of a single word recommendation (either UP or DOWN). Stock Analyst A has accuracy of 80%, stock analyst B has accuracy of 60%. Suppose today both analysts is saying UP, and assuming the two analyst reports are independent, what is the probability that apple stock will go UP? (if you don’t know how to compute this, for partial credit, please make a guess at the answer)
2. Continuing from Q1 Stock Analyst B is upset that Stock analyst A has a better record, but he realize he can improve his accuracy by the following strategy. He waits for Stock Analyst A’s report to come out. If the recommendation disagrees with his recommendation then with 0.7 probability he will publish the same recommendation as Stock Analyst A and with 0.3 probability he will publish his own report. Suppose today both analysts are saying UP, what is the probability that apple stock will go UP? (if you don’t know how to compute this, for partial credit, please state if the answer to Q2 is larger or smaller than the answer to Q1) BONUS: what is Analyst B’s new accuracy?
3. Continuing from Q2/Q1 Stock Analyst B’s managers are becoming increasingly suspicious that Stock Analyst B is no longer doing his own research. The suspicion is that maybe Stock Analyst B simply waits for Analyst A’s report to come out and with probability *p* he agree with analyst A and with probability *(1-p)* he disagree with Analyst A’s recommendation. You are hired as an external consultant to test (statistically) if this suspicion is true: you are given lots of historical data on Analyst A and Analyst B’s recommendations and Apple stock outcomes (UP/DOWN). How would you test this suspicion using a statistical method and present your finding to stock analyst B’s managers?
4. A professor receives performance & profile data for every employee of a large company. Upon a quick ANOVA analysis, the economist found that employees with criminal history perform 40% better than employees without criminal history. After adding a number of controls such as age, work experience, etc., the finding still holds. The economist believes that maybe because criminals need to plan for their clandestine activities, they are on average smarter than normal human beings. Can you help the professor identify a fundamental issue with the data analysis before he publishes a potentially flawed paper?
5. The King of Atlantis would like greater insights about his people so he can rule better. In particular, he would like to classify all of his subjects into either *good* or *bad*, and he has entrusted you, his chief Data Scientist with this task. The King’s previous data scientist had carefully sampled 1927 people from Atlantis and has given each of them a very accurate morality score. You can find the scores in the Google Fusion Table below

<https://www.google.com/fusiontables/DataSource?docid=1BEV071ATQi9IdNcJ0ZZMAIaL_w0hpIu2kz-ytOhP>

A positive score indicates a good person, and a negative score indicates a bad person. To comply with the King’s data protection laws, each person is anonymized and identified only by an unique string id.

Alas, the king has no patience and if you do not produce a classification result soon, you will be sent to the dungeon where the previous data scientist is now residing.

Luckily for you, the royal Scribe of Atlantis keeps a ledger of the social network at Atlantis. He has made available to you a list of friends as well as a list of foes at Atlantis via Google Fusion Table

Friends: <https://www.google.com/fusiontables/DataSource?docid=1aAoDt3wOI2A6cJe8YXvqptUohhyKceLeDqnyTBH2>

Foes:

<https://www.google.com/fusiontables/DataSource?docid=1RUDGQ5dQwvq8paZW9z1Cu6y9lyd6EfHiQvQgu-z6>

However, the royal Scribe’s apprentices are very sloppy and some of the friendships written on the ledger might not be real friendships. You can assume that this happens rarely (< 9% of the case). Most of the foes are documented correctly, though <1% of them might be false resulting from mistyping.

You may assume the following facts about friend/foe relationships

1. It is quite unlikely for a good person to be friends with a bad person
2. It is extremely unlikely for a friend of a friend to be a foe
3. It is extremely unlikely for a friend of a foe to be a friend
4. It is extremely likely for a foe of a good person to be good
5. It is very likely a foe of a bad person is good and a foe of a good person is bad

You task is to produce (1) 8 friendships in the friendships table that are most likely false (2) a classification of every individual you see (both from the friends table or the foes table) into either *good*, or *bad.* Please document your methodology and any further assumptions.