Call Centre

# Setup

This is a test on basic data processing. The test is tool-agnostic; you may use any data processing tool you like. Typical examples would include, Excel, python / pandas, matlab, SQL etc. The test is timed to last approximately 2 hours. Points available for each question are marked in [brackets].

You have been given 3 csv files:

1. leads.csv. This is a list of fictitious company directors, with some basic data about them and their company.
2. calls.csv. This is a list of fictitious calls made by an outbound call centre. The call centre consists of several agents, who make calls one after the other. They don’t get to choose who to call, the system does. The objective of the call is to get the lead to signup on the website. When they finish a call, they mark down the outcome, from a fixed list of possible outcomes. Note that a single lead may be called multiple times.
3. signups.csv. This is a list of leads who signed up after being called by someone from the call centre. Each signup was risk assessed and either approved or rejected for a loan.

# Questions

1. Which agent made the most calls? [1]
2. For the leads that received one or more calls, how many calls were received on average? [2]
3. For the leads that signed up, how many calls were received, on average? [2]
4. Which agent had the most signups? Which assumptions did you make? (note that there is a many-to-one relationship between calls and leads) [4]
5. Which agent had the most signups per call? [2]
6. Was the variation between the agents’ signups-per-call statistically significant? Why? [5]
7. A lead from which region is most likely to be “interested” in the product? [3]
8. A lead from which sector is most likely to be “interested” in the product? [1]
9. Given a lead has already expressed interest and signed up,
   1. signups from which region are most likely to be approved? [2]
   2. Is this statistically significant? Why? [5]
10. Suppose you wanted to pick the 1000 leads most likely to sign up (who have not been called so far), based only on age, sector and region.
    1. What criteria would you use to pick those leads? [10]
    2. In what sense are those an optimal criteria set? [3]
    3. How many signups would you expect to get based on those called leads, assuming they were being called by random agents? [3]
    4. If you could choose the agents to make those calls, who would you choose? Why? [3]