

Growth Analysis Task

Data Driven Insights

This task focuses on assessing your ability to tease insight out of data. It is assumed you will use Excel or Google Sheets, although you may use any tool at your disposal. There is no fixed time limit for the test – send back the results when you are happy you have had a good shot.

You are tasked with analysing results from an outbound call centre where agents are making calls to encourage people to sign up for a business loan.

You have been given 3 CSV files with data:

- **Leads.csv:** This is a list of contact details that were fed into the call centre's systems for agents to call (one row per person). Lead names and phone numbers are unique.
- **Calls.csv:** This is a list of calls made by the call centre to leads (one row per call). Calls are numbered from 1 to 9999 in chronological order. There may be no calls or multiple calls to some leads. Each call is given an outcome – for instance 'interested'. Note that a lead with an 'interested' call has not necessarily signed up and one lead may have multiple 'interested' calls.
- **Signups.csv:** This is a list of leads who actually signed up (this may or may not have been after an 'interested' call). Each signup was risk assessed and either approved or rejected for a loan.

Based on the data sheets, please answer the following questions. Show as much of your approach as possible – you will get credit for incorrect answers where you used a correct method.

1. How many calls did each agent make?
2. How many signups did each agent get? Assume a signup is assigned to the last agent who called the corresponding lead, in cases where multiple agents called the lead. Remember signups are marked down in signups.csv (a lead having an 'interested' call does not mean they signed up).
3. Which agent had the most signups per call? How many?
4. What was the median & mean number of calls to a lead before that lead was first marked as interested? Which is most appropriate in this situation?

5. A lead from which region was most likely to express interest, assuming they received at least one call? What proportion of leads from this region expressed interest? Assume that a lead with at least one 'interested' call has expressed interest.
6. Signups from which region are most likely to be approved? What proportion get approved?
7. What are the approximate error bounds of the proportion calculated in (6). Describe any assumptions.

IMPORTANT

Remember, part of this task assesses you on how well you present the data, which visualisations you choose and how easy your analytical process is to follow. **For example a great answer would look something like this:**

- a) How many calls with a call outcome = 'INTERESTED' did Agent Black make?

I used a pivot table, filtered on Agent Black displaying a count of the rows in the calls.csv dataset. **Agent Black made 365 calls marked as 'INTERESTED'.**

I used a bar chart to visualise this data as below:

