# Data Engineer Challenge - English

BairesDev specializes in delivering technology solutions. One of our engagement models is IT Staff Augmentation, where our engineers are an extension of the client’s teams. Most of our developers (approx 95%) are located in LATAM

We have developed an email marketing application to expand our client’s portfolio. The main responsibilities of the system are extracting data and sending email campaigns to potential customers.

We are attaching two folders containing:

* Contacts: Information about the leads that we have contacted in the past. It includes a file called *Contacts\_Positive* with information about leads that replied interested and several *ContactsN* files with mostly no replies or not interested contacts. The file has a column *score* that represents the business value that our sales team assigned to the potential customer, based on their profile (whether he is a decision-maker or not) and the potential of the opportunity (whether he is requesting one resource or a team of 20 developers).
* Data: Leads that we are planning to contact (they may have been contacted in the past)

We request you to analyze the data and try to predict which people from the *Data* files have the highest chances of increasing the business value if contacted. The expected output is at least a document explaining the analysis and a sheet including the rows of the potential leads we should send an email to.

* We tried to provide you with enough data to draw some conclusions. However, if you think the provided sample is not statistically significant or is missing key information, please let us know :-)
* Which additional data do you think would be relevant to improve the algorithm?