Capstone Batch 7 - client briefing

Email from your head of engineering

Dear new employee,

Welcome to the team, and congratulations on coming through the other side of the interview process! We're glad to have you as the newest member of the *Awkward Problem Solutions™* team*.*

As I'm sure you know from the interview process, we are a consultancy that tackles the hard data science problems no-one else will touch. We already have your first assignment. Don't let us down.

The Florida Department of Corrections has hired us to help address a sensible issue. Some years ago, they hired another company to design and implement an algorithm to predict whether a given individual would commit another crime in the future. However, there is growing concern that the algorithm has been unfair towards some ethnicities and backgrounds. Considering the high rates of incarceration, it is important that such algorithms are regularly assessed regarding fairness and, if necessary, adjusted or altogether replaced.

They have hired us to investigate whether the data shows any bias or fairness issue. If we are successful in investigating, they want us to create and host an API endpoint to predict recidivism. This will be integrated into a system to be used by judges and probation and parole officers to assess criminal defendants' likelihood of becoming recidivists. You will be responsible for setting this system up, and running it for a year.

I have attached the email from Dr. Penelope Dyson, the head of IT from the Florida Department of Corrections. Dr. Dyson will be your primary point of contact during this project. You must ensure that you disambiguate all requirements, as I believe the project may still be under-scoped in some aspects.

As you may imagine, this project is a huge deal for our consultancy, we're counting on you to deliver on this one. Naturally if you have any questions you can ask me, but I'm counting on you to lead and execute from start to finish.

Yours,
Henry
VP of Data Science
Awkward Problem Solutions™.

Email from Dr. Penelope Dyson

Dear Henry

I'm happy to hear that we'll be working together again. As you know we have been having some unfortunate coverage in the press regarding our recidivism prediction algorithm. I'm unaware of how credible these claims are, and want you as an external party to investigate and present an objective proposal and report.

Additionally we want your company to design and implement a system that will be used at courts to predict whether an individual will commit a crime in the future. It will be integrated into our own internal system, so all we need is a REST API that our own code can call.

Regarding data, we've been collecting information on criminal defendants, namely:
- Personal data (name, sex, age, race)
- Case data (e.g. in and out jail dates, charge degree, number of prior crimes committed)
- Data on previous arrests
- Recidivism scores for general criminal activity and for violent crimes
- Whether they reoffended or not

The **training dataset** has approximately 6,000 observations, I expect this should be enough.

Regarding the analysis, we want you to search for evidence that there is any bias on the previous recidivism scores (for general and violent criminal activity) towards an ethnicity or socio demographic group and whether there are any noticeable trends for specific criminal charge categories.

As you may imagine these are very delicate topics, and we want to ensure that we treat both the public and our officials with respect. We therefore need you to propose what metrics would suggest that anything unacceptable is being observed.

Naturally there may be trends, so if something seems to have already been corrected it is less relevant than if it still occurs.

Our expectation from experience is that some populations exhibit higher levels of delinquency than others, but would expect that the recidivism prediction would not vary significantly between populations.

The proposal and report will be read both by me and by members of our policy committee, please use this structure, and ensure that each section is written for the relevant audience.

After the analysis is completed and your report has been understood we will run a proof of concept with your company for the use of your API for a few of our courts.

The main objective of this API is to predict whether a criminal defendant has high risk of reoffending or not, i.e. to replace the previous recidivism prediction algorithm for general criminal activity..

We expect your API to be able to level the prediction between ethnicities and socio demographic groups, as well as between criminal charge categories.

Some data has been removed from the dataset, to be used in API test rounds.

From your last email I understand you've hired a specialist to deal with this project. I'm very much looking forward to working with them, and am of course available to answer their questions as they occur.

Best regards
Dr. Penelope Dyson
Head of IT
Florida Department of Corrections