Capstone Batch 4 - client briefing
Mail from your head of engineering

Dear new employee

Welcome to the team, and congratulations on passing the grueling interview process! We're glad to have you as the newest member of our team here at *Awkward Problem Solutions™.*

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.

o

The United Kingdom Department of Police has hired us to help solve some delicate problems. They have a stop and search policy for every police department, that's supposed to ensure that officers stop people and cars when there is probable cause. There have been accusations in the press that the police tend to stop and search certain minorities at a higher rate than others.

Another sensitive area is related to the question of asking suspects to remove articles of clothing for search, in which the accusations are that women of certain age groups and ethnicities are searched more than others. I do not need to stress that these accusations are extremely serious, and can lead to the dismissal of officers.

They have hired us to investigate whether the data proves any of these claims. If we are successful in investigating, they want us to create and host an API endpoint for authorizing searches. This will be integrated into a police system that the officer will need to use, ensuring uniformity of decisions. You will be responsible for setting this system up, and running it for a year.

I have attached the email from Dr. Anabelle Wilson, the head of IT from the Department of Police. Dr. Wilson 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. Annabelle Wilson

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 stop and search operations. I'm unaware of how credible these claims are, and want you as an external party to investigate and present an objective report.

Additionally we want your company to design and implement a system that our officers will need to use to approve the stopping of a person/car. It will be integrated into our own internal approval system, so all we need is a REST Api that our own code can call.

Regarding data, we've been collecting information on the search operations since 2017, with the following:

- Age range
- Date
- Gender
- Latitude
- Legislation
- Longitude
- Object of search
- Officer-defined ethnicity
- Outcome
- Outcome linked to object of search
- Part of a policing operation
- Removal of more than just outer clothing
- Self-defined ethnicity
- Type
- Station

The **training dataset** has approximately 660,000 observations, I expect this should be enough.

Regarding the analysis, we want you to search for evidence that any of our stations may be discriminating on gender, ethnicity or age regarding who they chose to stop, and also who they ask for any clothing to be removed.

As you may imagine these are very delicate topics, and we want to ensure that we treat both the public and our officers 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. Some training and changes of policy may impact the data over time.

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

The 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 police stations.

The main objective with this API is that searches are performed only when there is more than 10% likelihood that the search will be successful. We will evaluate this per station and per search objective, as we believe that some stations over search on certain search types.

We also expect that your API should be able to level the discovery rate without significantly diminishing our overall ability to detect offences. More explicitly, we want to level the discovery rate between ethnicities for every station and for every search objective.

The data from the year 2020 will not be in the training set, and we will call it as if it is live with data from that year. We won't be sending you any requests from searches we didn't perform, as naturally we wouldn't know the outcome.

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. Annabelle Wilson
Head of IT
Department of Police, Her Majesty's Government, UK