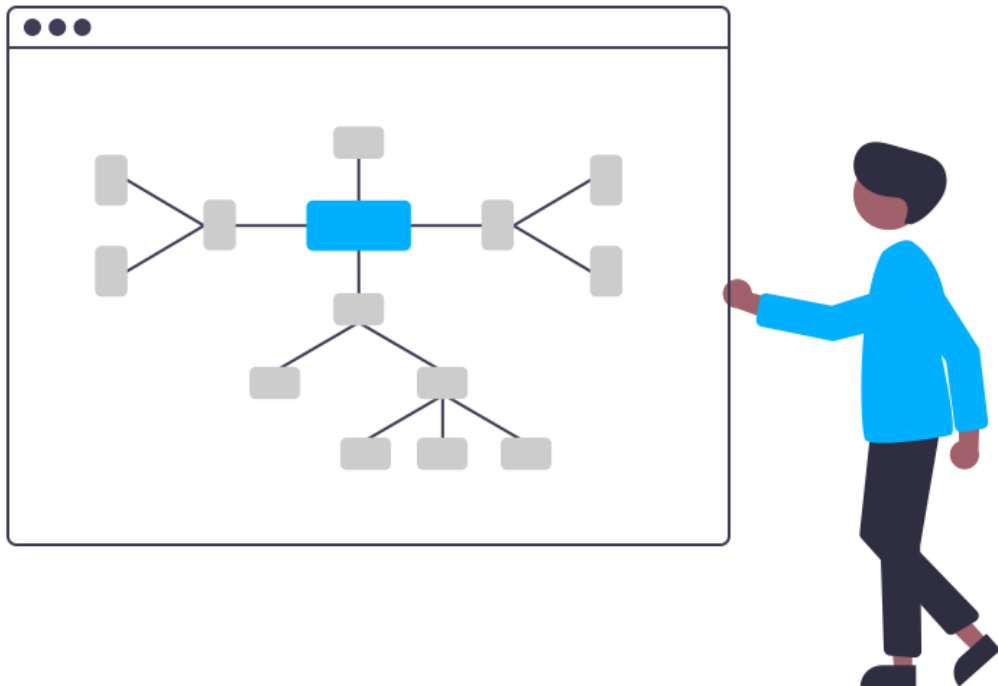# 2022

Data Engineer Interview Task

Data Engineer

Introduction

Congratulations on progressing to the second and final stage of the interview process for this role.

Prior to the interview, we ask that you complete the following data engineering exercise, which is designed to replicate the kind of challenge we are set by our clients and give you a flavour of the kind of work you would undertake as part of your role at ADSP.

In the interview, we'll ask you to take us through your approach and will ask you some technical questions relating to the challenge. Also have a think about how your work could be deployed to a live environment if required and the challenges you may face if doing so.

Good luck!

David Foster
Partner
Applied Data Science Partners (ADSP)

The Task

The Metropolitan Police Service are looking to understand how their stop and search policies have changed over time.

The latest data is available through the public Police Data API
https://data.police.uk/docs/method/stops-force/

Your task is to complete following steps:

- Write a script in Python to programmatically pull down all stop and search data from the API for the Metropolitan Police Service.
- Combine all data into a Pandas dataframe.
- Clean and format the data as you see appropriate.
- Write the data out to a csv file.
- Write a process that checks to see if new data has been added to the API at a given time each day - if so it should pull the new data and join it to the existing csv file.
- Prepare a short explanation of how you would store the data in a relational database if required – think about the structure of the data and what schema you would apply.
- Suppose entries were updated or deleted in the API – how would your propose that your system should handle this?
- Prepare an example of how you would create a view on the database to return an aggregate count of the number of each crime type by date.

Other things to consider:
- If there are other technologies that you want to include in your solution, do feel free – don't feel constrained by the brief. This is your chance to show off all your data engineering skills!

Good luck and have fun!