Technical Assessment

This part of the interview requires you to undertake a data analysis assignment. You can approach this problem and define your own analysis as you see relevant, and you will be expected to present your approach and relevant findings during the interview for 15 - 20 minutes. The assignment is completely open and as such there is no solution for which you are searching, this is about how you approach the analysis of a new dataset. We would prefer the analysis to be conducted using a suitable programming language, such as Python, R or SQL, using any tools of your choice.

During the interview you will be expected to walk us through your approach to analysing this data and to evidence any of your analysis with the relevant code in your chosen language. Example topics for discussion will be for you to explain why you made certain decisions, why you felt that was an appropriate approach/set of assumptions, what questions you were trying to answer, what challenges you encountered, etc. We are interested in how you develop hypotheses, conduct analysis and your ability to communicate results.

The data for analysis is hosted on the nycopendata.socrata.com website and is entitled "311 Service Requests from 2010 to Present". The available data spans over 10 years of requests for access to non-emergency municipal services in NYC. It is over 10GB in size, has 41 columns and contains over 25M rows of data. Additional information, data samples, glossary etc. are all provided on the same site.

The data can be accessed here: https://nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-toPresent/erm2-nwe9

Note that you are not limited to only this data, if you feel there are additional data sets that would enhance the analysis these can be incorporated as you require.