

Candidate Handout – Data Engineering Technical Case Study



Confidential and proprietary: Any use of this material without specific permission of McKinsey & Company is strictly prohibited

Problem Statement

A Formula 1 race team have approached QB with the aim of improving the efficiency of their production process.

- The project is still in it's 'Intelligence'(Beginner) phase where we're working how exactly we're going to go about solving the problem.
- Our initial hypotheses focus around employee productivity, equipment breakdowns and mistakes made during the manufacturing process.
- The first cut of the data has just arrived, these extracts have come from two parts of the client's organisation who rarely interact with each other.
- Initially the client has asked for some high level opinions and descriptive statistics from the data they have just shared.
- On the QB side, you have also been working closely with the Data Scientists and the Engagement manager on the study who are keen to start building variables which will go into the predictive model.
- The questions you will be asked in the interview partly feedback to the client, but also lay the foundations for the Data Science process.

Time allowed: 60minutes

This is an opportunity to show case your technical and problem-solving skills when faced with ambiguity.

During the interview, you will be provided with several structured data sources such as CSV, XML or JSON files that we would like you to interrogate:

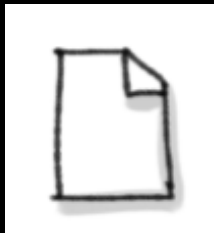
- You will also be asked to complete some questions and to explain your methodology, and
- At the end of the case study, you will be asked to send the code and output to the interviewers for further review

Environment for the case study:

- As this will be a coding interview you will need to have the right tools installed LOCALLY on your own laptop.
- We would like you to conduct your analysis using a Python notebook tool or another tool of your preference
- If you are going to use online tools from clouds such as AWS, Azure or Google, ensure you also have the local tools installed such as Python, Spark and a Jupyter notebook in case there are connection issues.

Preparing for Case Study Interview: Setup for success

1. **Make sure you have access to a laptop and charger for at least 2 hours with internet access**
2. **Make sure you have necessary development environment setup on your laptop to work with raw CSV or Excel files, for example one of the following:**
 - Python (Jupyter Notebooks, Pandas etc.)
 - Spark (Scala or PySpark)
 - Java



Python

