Technical Assessment

Data Analyst

Materials provided for the assessment:

- Dataset can be found here: http://archive.ics.uci.edu/ml/datasets/Online+Retail#
- This guidance document for the technical assessment

Expected outputs:

- Python code or Python notebook
  - Preferably hosted on repl.it https://repl.it/ (create a login and then send over link to code)
  - Alternatively, Github public repo, link to be sent over
  - Last recourse option: send over Python code or Python notebook by email.
- Where analysis wants to be performed in Excel, Excel spreadsheet sent over by email, however, please be cognisant of volume of the dataset.
- A document, format to be determined by the candidate, answering the below questions.

Criteria for evaluation:

- It is expected that the code can be run with no assistance (so ensure requirements to run are described in associated documentation).
- It is also expected that the document hosting the responses to the below questions can also be opened and browsed with no assistance.
- Quality of exploratory data analysis.
- Uncovering of trends in the data.
- Communication of results and decisions being made in the process.
- Quality of visualisations.

You have 1 week to submit results back to complete this technical assessment.

Practical questions:

Question 1:

At first glance, what would be your comments regarding this dataset and its representativeness of real-world datasets? Please elaborate.

Question 2:

Please provide some commentary to the exploratory data analysis to be performed on the dataset. What has the exploratory data analysis allowed you to uncover?

Question 3:

Now that you have performed some exploratory analysis of the dataset, would you amend the dataset in any way, and how will you ensure traceability of the modifications or enrichments?

Question 4:

What questions would you want to answer with this dataset, and how would you go about setting up the analysis for them? Please provide an example of at least one analysis on this dataset, with explanation and visualisations.

Question 5:

Please conclude your analysis with a summary of key findings.

Question 6:

What skillsets do you have that you feel have not been highlighted by this technical assessment?

Theoretical questions:

Question 7:

When defining key performance indicators and metrics, what is the approach you take for effective design?

Question 8:

When defining metrics for validation of a linear regression, what is the approach you take for effective design? What are the pros and cons of e.g. MSE or MAE?

Question 9:

When performing a linear regression, what are the inherent assumptions that are made on the dataset? What are the drawbacks of a linear model, and when can it not apply?

Question 10:

The data and analytics engineering team at Laing O'Rourke will be built on 3 profiles: data analysts, data scientists and data architects. Please provide your understanding of how those roles fit together and complement one another, highlighting any examples you may have come across at any scale.