Bank of England

Technical Assessment – Data Scientist - 010030

A Supervision Manager has asked you to help allocate scarce resources and identify which firms their team should focus on. Supervisory resource may be allocated according to the following characteristics:

- Firm size (i.e. the biggest firms need more attention)
- Changing business profile (are firms' data changing substantially year-on-year?)
- Outliers from the norm (when looking at a single reporting period, does a firm deviate significantly from the average?)

Some typical metrics have been provided in the attached data sheets. These include:

- Gross Written Premium (GWP) – total revenue written by an insurer. Equivalent of turnover for a non-insurance firm.
- Net Written Premium (NWP) – GWP less reinsurance. NWP / GWP will show how much of the firm's risk is being passed on to reinsurers.
- SCR coverage ratio – a measure of whether a firm is meeting its prudential capital requirements. Greater than 100% means the firm is holding enough capital to meet the requirement. The size of the buffer (i.e. surplus over 100%) can be important.
- Gross claims incurred – a large cost to an insurer. Monitoring how these change over time for a firm is vital.
- Net combined ratio – (incurred losses plus expenses) / earned premiums. This is a ratio that can indicate the profitability of a firm. If this is less than 100% it indicates a profit.

Task I

Using the data provided, please analyse this data using R or Python and produce a short report (maximum 3 pages including tables and charts), to highlight which firms should receive the most supervisory attention, according to the metrics above. In your report, please include reflections on any processes undertaken to clean the data, the methodology behind the decisions why certain firms should receive more attention, and what tools or techniques have been chosen for these decisions.

Task II

Given the data provided, implement relevant ML techniques in R or Python to draw out further insights. Please reflect on relevant anomaly detection algorithms, appropriate ways to measure their performance, and data cleaning processes you may have undertaken. Please also include your thoughts on automating your completed work in the context of continued analysis. Present your insights as a separate section in your report (maximum 3 pages including tables and charts).

Notes

- More metrics have been provided in the data than are necessary to include. Feel free to utilise these as you wish, but it is not necessary to use all metrics.
- The data have been anonymised using a random multiplier. As such, there may be some unexpected patterns in the data.
- It can be useful to include multiple metrics on the same chart e.g. net combined ratio on one axis, net written premium on another. Joining of the datasets may be required to achieve this.
- Some of the data will contain outliers that may be genuine, but some may be errors in reporting (i.e. numbers that are so different to all other firms they are obviously wrong). Try to distinguish between the two, a separate section highlighting errors in reporting is also useful.
- Should you pass this assessment stage, you will be asked to explain sections of your code, as well as your general approach to the analysis.