Data science exercise

This exercise aims to help you showcase your proficiency in data science by utilizing property sales data.

The Data

HM Land Registry publish the "Price Paid Data", detailed data on the sale prices of properties in England and Wales. Datasets are published monthly with transaction-level information, including price, date, location, property type, property age, tenure duration, and transaction type. Datasets are available in bulk in multiple file formats and filtered via the Linked Data API. Please use whatever source is most appropriate to your analysis, acknowledging it appropriately in your report. More information can be obtained from the Gov.uk Price Paid Data website.

What you need to do

Please, analyse property price data to gain insight into the housing market. Feel free to define your focus and to narrow down the scope of your work. Include statements on how the housing market has changed since the Covid-19 pandemic, along with other compelling and statistically rigorous insights. Present descriptive statistics, visualisations, and machine learning models. Please define the scope of your analysis based on your interests, experience, and time constraints. Remember that the purpose of this exercise is not to conduct exhaustive analysis, but to demonstrate the range of skills that this role requires against the assessment criteria.

Outputs

You should provide:

A. Well-written code that generates the results, written in a language of your choice.
B. A write-up of your analysis and results, in the form of a briefing written to policymakers who do not have a technical background in data science.

The write-up should be no longer than 3 pages or slides including footnotes, tables, charts, appendices etc. It should contain the following information (presented in an order of your choosing):

1. An overview of the dataset you chose to use and the steps you took to prepare it.
2. An effective visualisation of some of your findings.
3. An explanation of your modelling approach and an assessment of model performance.
4. A summary of the results.
5. Suggestions for further analytical work and policy recommendations.
6. An identification of challenges from performing your analysis at scale, and how you might address these.

Assessment

Please email the write-up and the code used to: rosie.smith@bankofengland.co.uk

If the deadline for the task stated in the cover email is not achievable due to other commitments, please let us know as soon as possible.

We will assess the exercise according to the following criteria:

- The quality of your code.
- The quality of your method choices and the discussion of their advantages and limitations.
- The clarity of your presentation, including text and visualisations.
- Your ability to link your findings to policy recommendations.