**HTX xData Technical Test Question – Data Scientist**

Instructions:

* You can make assumptions to any information not specified within the Test Instruction.
* Create a public repository for submission of deliverables. For each task group (Model Prediction and Link Analysis), please provide all code, output, and answers in a single Python notebook. Proper code documentation, appropriate markdown formatting for written answers, and accompanying visualisations are expected. Ensure your explanations are clear and accessible to both technical and non-technical audiences.
* Please complete the tasks and email the repository link to the sender **within 7 calendar days** upon receiving this email.
* Reach out to us if you have any queries via email.
* All submitted artifacts must be executable without errors. If a specific environment setup is required, please include the setup scripts in the submission. Failure to comply will result in the submission being considered incomplete.

Model Prediction Task I

Consider the following scenario:

You are tasked to build a predictive algorithm to determine the factors affecting prices of residential properties in Singapore. You need to provide insights to your reporting officer to detail one or more strategies in curbing housing prices inflation.

Your fellow colleagues should be able to access and contribute to your code to replicate the same insights. Their local devices do not have GPU access. Provide justification for any of the choices you have made.

Datasets:



Model Prediction Task II

Explain the factors and considerations in building an in-house predictive model for users. You may use the above scenario to support your answer with reasonable assumptions.

Link Analysis Task I

Based on the data (Jan 2017 onwards) given in Model Prediction Task I, create a graph to analyse relationships between each property listing. You are free to choose any graph schema you prefer, as well as feature construction and graph construction methods, with adequate justification. Include high-level analysis of the graph components.

Link Analysis Task II

On the graph created in Link Analysis Task I, run one or more graph algorithms of your choice. For each algorithm, explain in your own words how the algorithm works, and provide a plausible interpretation of the algorithm output.

Bonus Question: Link Analysis Task III

Note: This is an optional question which good attempt on this question will gain additional score.

On the graph created in Link Analysis Task I, train a graph neural network for housing price prediction. You are free to choose your own model architecture, loss function, and appropriate data transformation and feature engineering techniques, with adequate justification. You are also free to choose your own data split ratio, as long as the results are reproducible. Explain the results of your model.

**Submission Instructions**

Create a public git repository e.g., <https://github.com/fred/myrepo>. You may use any of the publicly available repositories like GitHub, Gitlab, etc. Repository should include the following files:

1. For each task group (Model Prediction and Link Analysis), please provide all code, output, and answers in a single **Python notebook**. Proper code documentation, appropriate markdown formatting for written answers, and accompanying visualisations with documentation in proper markdown formatting are expected.
2. A **requirements.txt** for all Python libraries and version required
3. Include a **README.md** file to document the steps for setup and instructions to run your code.
4. **Slides** to explain your work. This will be used during the interview, should you be selected, to explain your solution. Slides should include explanation and logic on:
   1. Assumptions
   2. Date exploration and preparation steps
   3. Models explored and discussion on the performance of models
   4. Findings on results
   5. Other further exploration that could be done