Question: Explain the factors and considerations in building an in-house predictive model for users. You may use this HDB example to support your answer, and please state any assumptions made.

1. Data Governance

- Factor: It is important to comply with whole of government (WOG) data regulations such as IM8. For example, storage, usage, retention and management of sensitive or personal data.
- Consideration: It is essential to ensure that the data analysis is conducted ethically, securely and in a manner that complies with IM8.
- HDB_Prices Example: When processing the HDB_Price dataset, it is necessary to ensure that it is not handled in a biased manner such as altering values to align with business objectives. It is also crucial to secure any sensitive or personal data during the process. After generating the predictions, establish a retention period and ensure that the data is stored in the designated storage location. All these procedures must in must be in accordance with IM8 regulations to ensure data handling.

2. Expertise

- Factor: Building a predictive model requires expertise in machine learning. It is not just about building the model but also fine tuning the model to improve predictions.
- Consideration: It will be necessary to hire professionals like data scientists, machine learning engineers and analysts to design and build the model. Hiring individuals with the right skills often results in higher costs and sometimes longer wait times to fill the vacancies.
- HDB_Prices Example: Building a predictive model requires proficiency in algorithms and a solid understanding of the theoretical aspects of machine learning. Proficiency in algorithms is essential to create the machine learning model and a strong knowledge of theoretical concepts is crucial for choosing and applying the correct model. As such, it is a specialized role and filling such positions often involves higher hiring costs and longer waiting times as not everyone possesses the required skill sets.

3. Cost (e.g., Hardware and GPU)

- Factor: Running more advanced models (e.g. deep learning), especially working with large datasets or complex algorithms like Support Vector Machine requires powerful hardware. This often means using better GPUs which can be expensive.
- Consideration: It is important to assess the costs of hardware such as upgrading to better laptops with more powerful GPUs that can handle large datasets and complex algorithms.
- HDB_Prices Example: Having a larger dataset such as one spanning from 1990 to 2020 or beyond requires a more powerful GPU to ensure machine learning models run efficiently and quickly. As the volume and complexity of the data increase especially dealing with time series data, GPU with greater processing power becomes essential. Without this, model training times can be significantly prolonged and the system might struggle to handle the data at scale. Upgrading GPU can be costly and may exceed the project budget. Therefore, it is necessary to assess both the costs of the upgrades and ensure they are aligned with the project budget while allowing the system to handle the dataset efficiently.

4. Regular Maintenance and Updates

- Factor: Once the predictive model is up and running, ongoing maintenance and updates are essential to keep it performing optimally and running smoothly.
- Consideration: It is crucial to monitor changes in the data. Failing to retrain or update the model with new data could lead to outdated predictions, poor performance and uninformed decision-making.
- HDB_Prices Example: When introducing a new flat type in HDB such as a white flat with no walls, the algorithm must be updated to include this new flat type in the flat-type dictionary. This ensures the system continues to run smoothly and enables stakeholders to make informed and data-driven decisions based on up-to-date and accurate price estimations.