Problem Definition

The objective of this task is to predict the price of an apartment transaction in 2017 and onward, based on the price information from previous years.

The data set provided includes information on housing trade from 2009 to early 2018, you are expected to split the data into training set and testing set, based on their tradeTime.

The data set including ID, Lng, Lat, CommunityID, TradeTime, DOM (days on market), Followers, Total price, Price, Square, Living Room, number of Drawing room, Kitchen and Bathroom, Building Type, Construction time, renovation condition, building structure, Ladder ratio, elevator, Property rights for five years, Subway, District, and Community average price. Here is the more detailed explanation for each of the column:

- ID: Transaction identifier
- Lng: Longitude of location
- Lat: Latitude of location
- Cid: Community identifier
- tradeTime: Transaction timestamp
- DOM: Active days on the market
- followers: Number of people watching this transaction
- square: Number of square meters of living space
- price: The average price per square meter
- livingRoom: Number of living rooms
- drawingRoom: Number of drawing rooms
- kitchen: Number of kitchens
- bathroom: Number of bathrooms
- floorPosition: 5 types
    - Top: Apartment occupies highest floors in the building
    - High: Apartment occupies floors above Middle and below Top
    - Middle: Apartment occupies floors in the Middle
    - Low: Apartment occupies floors below Middle and above Bottom
    - Bottom: Apartment occupies lowest floors in the building
- floorsCount: The total number of floors in the building
- constructionTime: The time of construction in years
- renovationCondition: 4 levels
    - 4: Best condition
    - 3: Good condition
    - 2: Rough condition
    - 1: Other
- buildingStructure: 6 types
    - 6: Steel and concrete
    - 5: Steel
    - 4: Concrete

    - 3: Brick and wood
    - 2: Mixed
    - 1: Unknown
- ladderRatio: The ratio of elevators to apartments on the same floor
- elevator: 2 types
    - 1: yes, the building has an elevator
    - 0: no, the building does not have an elevator
- fiveYearsProperty: 2 types (this sounds counterintuitive, but the types are confirmed)
    - 1: yes, the owner has had the property for less than 5 years
    - 0: no, the owner has had the property for more than 5 years
- subway: 2 types
    - 1: yes, there is a subway nearby
    - 0: no, there is not a subway nearby
- district: The numerical district the building is located in
- communityAverage: Average price of the transaction done in this community

Technical Requirements

- The solution should be developed in Python. Using Jupyter notebook or other similar tools that can record the data exploration process is required.
- The use of data mining/machine learning algorithms and libraries, such as Scikit-Learn, is encouraged.
- The solution should be able to handle missing data, outliers, and imbalanced data.

Evaluation Criteria

The solution will be evaluated based on the content in your notebook. The thinking process/ data exploration process is equally important as the end result. The solution's efficiency in terms of processing time and memory usage will not be considered.