The following are the data cleaning processes we followed.

1. Basic Cleaning of Data

* Dropping all Arabic columns
* Meter sales price >=1 and upper bound (Q3+ 1.5IQR)
* Procedure area >=1 and upper bound (Q3 + 1.5IQR)

2. In transaction data for column ‘Property\_type\_en’

* for main transaction\_2020.csv data - Select year >=2020.[**(transaction\_2020.csv)**](https://drive.google.com/file/d/10dxLK7-TqruaeMKaoUQs6VKo7rCRsQ55/view?usp=drive_link)
* Units in transaction.csv as units\_2020.csv - Select year >= 2020.[**(units\_2020.csv)**](https://drive.google.com/file/d/1-1ECd71N1X2AirUUK-vfJtdCJLPNNKGt/view?usp=drive_link)
* Villas in transaction.csv as villa\_2020.csv - Select year >=2020. [**(villas\_2020.csv)**](https://drive.google.com/file/d/1-4rBS2c_8XsgWXOyRvcQqf5a-OZrYFNe/view?usp=drive_link)
* Land in transaction.csv as land\_2020.csv - Select year>=2020. [**(land\_2020.csv)**](https://drive.google.com/file/d/1-81wGSmnfJdVGW3RjF13I6EVZAkUvncT/view?usp=drive_link)
* Building in transaction.csv as building\_2020.csv - Select year>=2020. [**(Building\_2020.csv)**](https://drive.google.com/file/d/1-BUVHKW_HybH-U2xOjAgJ4NWDWGDhBvS/view?usp=drive_link)

3. Creating dummies

* Created dummies for all the data
* Created dummies for the columns mentioned in doc secondary variables.

4. Running Model