XI LI
SIU KI WONG
CHURONG WANG

VALUATION OF HONG KONG RESIDENTIAL PROPERTY

Dataset Description

The data set 'Centaline_data.csv' contains information on more than 180,000 property transaction records in Hong Kong between 2018-2021. Each row indicates one property transaction record and includes more than 50 variables, such as the number of rooms, gross size, and transaction price of the property. Most outliers were removed from the dataset. The data set is randomly split into two pieces — a training set called 'Centaline_train.csv' (with 140,000+ transaction records) and a testing set called 'Centaline_test.csv'(with 40,000+ transaction records). The entire set, the training set, or the test set could be used on different occasions. A shapefile of the Hong Kong map is also provided for geographic visualization.

Variable Information

The data has the following features, with Transaction_price being the target (dependent) variable.

Name	Data Type	Measurement	Description	Range
Transaction_price	numeric	Hong Kong Dollar (HKD)	The transaction price of the property	400,000-639,796,000
Transaction_year	numeric	Year	Transaction year of the property	2018-2021
Transaction_month	numeric	Month	Transaction month of the property	1-12

Name	Data Type	Measurement	Description	Range
Location	string		Location of the property	
Estate	string		The estate that the property is in	
HMA	categorical		Housing Market Area	161 variables
Developer	categorical		10 biggest developers in Hong Kong and "Other"	11
Gross_size	numeric	Square foot	Saleable area plus a share of the common areas, bay windows, and facilities	100-6,158; -1 means no data
Saleable_size	numeric	Square foot	The actual size of a property, including balconies	91-4,616
No_of_rooms	numeric		Number of rooms in the property	0-6; -1 means no data; 0 means the property is a studio
Floor	numeric	Floor	The floor of the property	1-87
Region	categorical		The region where the property is located in	3 variables
Primary_school	categorical		Primary school Net divides Hong Kong's primary schools into 36 zones	36 variables
Secondary_school	categorical		Secondary schools use a zoning system based on the 18 districts in Hong Kong	18 variables
Age_of_property	numeric	Year	Age of the property since construction	0-71; -1 means the property was uncompleted by the time of the transaction
Uncompleted (binary columns)	categorical		Whether this property is completed or not by	0 – no; 1 – yes

Name	Data Type	Measurement	Description	Range
			the time of the transaction	
MTR_station	categorical		Mass Transit Railway (MTR) station that is close to the property	93 variables; -1 means no MTR station is close to the property
Close_to_MTR (binary columns)	categorical		Whether the property is close to an MTR station	0 – no; 1 – yes
Shopping_mall - Ballroom (binary columns)	categorical		Whether the property is equipped with the amenity	0 – no; 1 – yes
District	categorical		Hong Kong has 18 districts	18 variables
Median_income	numeric	Hong Kong Dollar (HKD)	The median income of residents in the HMA	12,210-187,500
Median_age	numeric	Year old	The median age of residents in the HMA	35-51.5
Population	numeric		Total number of people in the HMA	1,116-285,862
Unit	numeric		No. of private residential units in the HMA	79-36,019
Lat	numeric	Degree	Latitude of the property	22.20455-22.54544
Lon	numeric	Degree	Longitude of the property	113.93906-114.27483
Distance_to_Central	numeric	Meter	Distance from the property to Central	316.7-30,275.4
Second_Hand (binary columns)	categorical		Whether the transaction is secondhand or not	0 – no; 1 – yes