

Handling Strategies

Variable Names	Data Quality Issue		Handling Strategy
DateofSales(dd/mm/yyyy)	Some dates failed the tests.		Update the incorrect date.
Address	High cardinalities.		Keep as is.
PostalCode	Missing values(80.67%)	Extract postal code from 'Address' column, or refill by 'N/A'.	
PostalCode	Some instances failed the tests.	Remove the incorrect postal code information.	
County	None.		Keep as is.
Price(€)	Outliers.		Remove the outliers.
NotFullMarketPrice	None.		Keep as is.
VATExclusive	Some instances failed the tests.		Update the incorrect information.
DescriptionofProperty	None.		Keep as is.
PropertySizeDescription	Categories overlap.	Combine the overlapped categories.	
PropertySizeDescription	Missing values(89.45%)	Keep as is.	

Descriptive Statistics for cleaned categorical features

	count	unique		top	freq	%missing
Address	9454	9440	12 CHURCH ST, HOWTH,,Dublin 13	2	0.000000	
PostalCode	9454	23	N/A	7563	0.000000	
County	9454	26	Dublin	2792	0.000000	
NotFullMarketPrice	9454	2	No	8983	0.000000	
VATExclusive	9454	2	No	7877	0.000000	
DescriptionofProperty	9454	2	Second-Hand Dwelling house /Apartment	7877	0.000000	
PropertySizeDescription	1019	3	greater than or equal to 38 sq metres and less...	745	89.221494	

Descriptive Statistics for cleaned continuous features

	count	mean	std	min	25%	50%	75%	max
Price(€)	9454.0	211410.339013	125294.849686	5963.0	115000.0	191094.5	285337.5	594714.0