Data set explained

locations.csv.zip

List of locations, we have 4 levels of locations (country -> province -> region -> area), and in this file you just has id with parentId.

Field Name	Example	Description
id	15	Location ID
parentId	2	Parent Location ID

types.csv.zip

List of estate types (house, apartment, etc), here we have 2 levels (so you have id and it's group name).

Field Name	Example	Description
id	10	Estate type ID
groupName	house	Group Name

trainListings.csv.zip & testListings.csv.zip

Listing is real estate property found on agency website. Listings descriptions include metadata, texts, and list of thumbnail ids, etc. Data split between training and test like this: ~80% are in training data, and the rest ~20% in test data set.

Field Name	Example	Description
id	123	Listing ID
sourceId	1	Agency ID
locationId	15	Location ID
typeId	10	Estate type ID
price	10000000	Sale/ rent price
rooms	13	Some agencies use this as a total number of rooms
bedrooms	5	Number of bedrooms

Field Name	Example	Description
bathrooms	3	Number of bathrooms
totalArea	500	Some agencies prefer to show full buildup area
livingArea	300	Some agencies prefer to show only living area
plotArea	2000	Area of a plot
terraceArea	50	Area of a terrace
title	Luxury estate with large, park-like garden in Llubí	Listing title, plain text
description	This estate was built on one floor in a typical Spanish hacienda style. It impresses with an extraordinary use of many antique materials with an eye for details. The huge ceilings of approx. 6 meters give the property a manorial character. … has its own water. The pool is 24m long.	Listing description, plain text
features	Fireplace, Terrace, Accessible for wheelchairs, Mountain view, Air conditioning, Swimming pool, Garden, Built-in kitchen	Listing features list, plain text, in some cases can be poorly saved, without separators between words, be careful.
latitude	39.5441	Listing coordinate, present in rear case. *Caution: in some cases can be coordinate of agency office.*
longitude	2.38938	
thumbnails	3966,3967,3968,3969,3970,3971, 3972	List of thumbnail IDs. Thumbnail path explained below (thumbnails.zip).

---

trainMatchedListings.csv.zip

here you have matched and unmatched pairs. So, if listing #3 is the same as listing #5, you will see here 2 pairs: 3,5,1 and 5,3,1.

Field Name	Example	Description
id1	3	First listing
id2	5	Second listing
	1	Match or not: 1 - yes 0 - no

thumbnails.zip

thumbnails for all listings (train and test). In listings descriptions you have list of ids for thumbnails. They are separated between multiple directories: directory name = ceil(thumbnailId / 10000).

Example:

Thumbnail id = **559216**
ceil(559216 / 10000) = **56**
Thumbnail path = **/56/559216.jpg**