

train.csv - (1503424, 18)

ColumnName	DataType	#Nulls
item_id	object	0
user_id	object	0
region	object	0
city	object	0
parent_category_name	object	0
category_name	object	0
param_1	object	61576
param_2	object	654542
param_3	object	862565
title	object	0
description	object	116276
price	float64	85362
item_seq_number	int64	0
activation_date	object	0
user_type	object	0
image	object	112588
image_top_1	float64	112588
deal_probability	float64	0

train.csv - Train data.

item_id - Ad id.
user_id - User id.
region - Ad region.
city - Ad city.
parent_category_name - Top level ad category as classified by Avito's ad model.
category_name - Fine grain ad category as classified by Avito's ad model.
param_1 - Optional parameter from Avito's ad model.
param_2 - Optional parameter from Avito's ad model.
param_3 - Optional parameter from Avito's ad model.
title - Ad title.
description - Ad description.
price - Ad price.
item_seq_number - Ad sequential number for user.
activation_date- Date ad was placed.
user_type - User type.
image - Id code of image. Ties to a jpg file in train_jpg. Not every ad has an image.
image_top_1 - Avito's classification code for the image.
deal_probability - The target variable. This is the likelihood that an ad actually sold something.
It's not possible to verify every transaction with certainty, so this column's value can be any float from zero to one.

test.csv - Test data. Same schema as the train data, minus deal_probability.
train_active.csv - Supplemental data from ads that were displayed during the same period as train.csv. Same schema as the train data, minus deal_probability.
test_active.csv - Supplemental data from ads that were displayed during the same period as test.csv. Same schema as the train data, minus deal_probability.
periods_train.csv - Supplemental data showing the dates when the ads from train_active.csv were activated and when they where displayed.
item_id - Ad id. Maps to an id in train_active.csv. IDs may show up multiple times in this file if the ad was renewed.
activation_date - Date the ad was placed.
date_from - First day the ad was displayed.
date_to - Last day the ad was displayed.
periods_test.csv - Supplemental data showing the dates when the ads from test_active.csv were activated and when they where displayed. Same schema as periods_train.csv, except that the item ids map to an ad in test_active.csv.

test.csv (508438, 17)

ColumnName	DataType	#Nulls
item_id	object	0
user_id	object	0
region	object	0
city	object	0
parent_category_name	object	0
category_name	object	0
param_1	object	22910
param_2	object	233229
param_3	object	306331
title	object	0
description	object	0
price	float64	30585
item_seq_number	int64	0
activation_date	object	0
user_type	object	0
image	object	42609
image_top_1	float64	42609

Avito

periods_train (16687412, 4)

ColumnName	DataType	#Nulls
item_id	object	0
activation_date	object	5790
date_from	object	0
date_to	object	0

Sample Submission (508438, 2)

ColumnName	DataType	#Nulls
item_id	object	0
deal_probability	float64	0

train_active (47920403, 14)

ColumnName	DataType
user_id	object
region	object
city	object
parent_category_na me	object
category_name	object
param_1	object
param_2	float64
param_3	float64
title	object
description	object
price	float64
item_seq_number	float64
activation_date	object
user_type	object

periods_test (13724922, 4)

ColumnName	DataType	#Nulls
item_id	object	0
activation_date	object	5043
date_from	object	0
date_to	object	0

test_active (45331702,14)

ColumnName	DataType
user_id	object
region	object
city	object
parent_category_na me	object
category_name	object
param_1	object
param_2	object
param_3	object
title	object
description	object
price	float64
item_seq_number	float64
activation_date	object
user_type	object

Train Images # - 1,390,836
Test Images # - 465,829