

Skill test: Senior (Applied) Data Scientist, Property Intelligence Tribe

Thank you for going through the first steps of the process, and congratulations on reaching the skill test stage. We understand it takes time, and we appreciate your effort. We're excited to see your skills in action!

What we do at PriceHubble

Automated Valuation Models (AVMs) are tools used in <u>real estate</u> to estimate a property's value on a particular date, without the need for an <u>appraiser</u>.

AVMs determine the value of a property for sale based on <u>Market Value</u>, which is the expected price in a fair deal between a willing buyer and seller.

Welcome to the team

Exciting news! We're growing and diving into the U.S. market, starting our journey in <u>Ames, Iowa</u>. As a Senior Data Scientist, your mission is to create the first MVP for our product launch.

Your objective: develop a model that predicts house sale prices using residential property sales recorded with the Ames Assessor's Office from 2006 to 2010.

The dataset (ames_transactions.csv) includes details on 2,930 transactions, covering:

- House specifics (bedrooms, garage, fireplace, pool, porch, etc.).
- Location details (neighborhood).
- Lot particulars (zoning, shape, size, etc.).
- Ratings for condition and quality.
- Sale price.

What we look for

We want good performance and accuracy, but we also need to explain what influences property value. It's also crucial to know when the model works well and when it doesn't, so we can target the right customers effectively.

Data Documentation

Name	Туре	Description
property_id	Nominal	Parcel identification number
dwelling_subclass	Nominal	Identifies the type of dwelling involved in the sale.
zoning_type	Nominal	Identifies the general zoning classification of the sale.
lot_frontage_ft	Continuous	Linear feet of street connected to property.
lot_area_sq_ft	Continuous	Lot size in square feet.
street_type	Nominal	Type of road access to property.
alley_type	Nominal	Type of alley access to property.
lot_shape	Ordinal	General shape of property.
land_contour	Nominal	Flatness of the property.
utilities	Ordinal	Type of utilities available.
lot_configuration	Nominal	Lot configuration.
land_slope	Ordinal	Slope of property.
neighborhood	Nominal	Physical locations within Ames city limits.
condition_1	Nominal	Proximity to various conditions.
condition_2	Nominal	Proximity to various conditions (if more than one is present).
dwelling_type	Nominal	Type of dwelling.
dwelling_style	Nominal	Style of dwelling.
overall_quality_rating	Ordinal	Rates the overall material and finish of the house.
		

overall_condition_rating	Ordinal	Rates the overall condition of the house.
year_built	Discrete	Original construction year.
year_remod_add	Discrete	Remodel year (same as construction date if no remodeling or additions).
roof_style	Nominal	Type of roof.
roof_material	Nominal	Roof material.
exterior_type_1	Nominal	Exterior covering on house.
exterior_type_2	Nominal	Exterior covering on house (if more than one material).
masonry_veneer_type	Nominal	Masonry veneer type.
masonry_veneer_area_sq_ft	Continuous	Masonry veneer area in square feet.
exterior_quality_rating	Ordinal	Evaluates the quality of the material on the exterior.
exterior_condition_rating	Ordinal	Evaluates the present condition of the material on the exterior.
foundation_type	Nominal	Type of foundation.
bsmt_quality_rating	Ordinal	Evaluates the height of the basement.
bsmt_condition_rating	Ordinal	Evaluates the general condition of the basement.
bsmt_exposure_rating	Ordinal	Refers to walkout or garden level walls.
bsmt_fin_type_1_rating	Ordinal	Rating of basement finished area.
bsmt_fin_type_1_sq_ft	Continuous	Type 1 finished square feet.
bsmt_fin_type_2_rating	Ordinal	Rating of basement finished area (if multiple types).
bsmt_fin_type_2_sq_ft	Continuous	Type 2 finished square feet.
bsmt_unf_sq_ft	Continuous	Unfinished square feet of basement area.
total_bsmt_sq_ft	Continuous	Total square feet of basement area.

heating_type	Nominal	Type of heating.
heating_quality_rating	Ordinal	Heating quality and condition.
has_central_air	Nominal	Central air conditioning.
electrical_system_type	Ordinal	Electrical system.
first_flr_sq_ft	Continuous	First Floor square feet.
second_flr_sq_ft	Continuous	Second floor square feet.
low_quality_fin_sq_ft	Continuous	Low quality finished square feet (all floors).
gr_living_area_sq_ft	Continuous	Above grade (ground) living area square feet.
bsmt_full_bath_count	Discrete	Basement full bathrooms.
bsmt_half_bath_count	Discrete	Basement half bathrooms.
full_bath_count	Discrete	Full bathrooms above grade.
half_bath_count	Discrete	Half baths above grade.
bedroom_abv_grd_count	Discrete	Bedrooms above grade (does NOT include basement bedrooms).
kitchen_abv_grd_count	Discrete	Kitchens above grade.
kitchen_quality_rating	Ordinal	Kitchen quality.
room_abv_grd_count	Discrete	Total rooms above grade (does not include bathrooms).
functionality_type	Ordinal	Home functionality (Assume typical unless deductions are warranted).
fireplace_count	Discrete	Number of fireplaces.
fireplace_quality_rating	Ordinal	Fireplace quality.
garage_type	Nominal	Garage location.
garage_finish_rating	Ordinal	Interior finish of the garage.
garage_car_count	Discrete	Size of garage in car capacity.
garage_area_sq_ft	Continuous	Size of garage in square feet.
garage_quality_rating	Ordinal	Garage quality.

garage_condition_rating	Ordinal	Garage condition.
paved_drive_type	Ordinal	Paved driveway.
wood_deck_sq_ft	Continuous	Wood deck area in square feet.
open_porch_sq_ft	Continuous	Open porch area in square feet.
enclosed_porch_sq_ft	Continuous	Enclosed porch area in square feet.
three_season_porch_sq_ft	Continuous	Three season porch area in square feet.
screen_porch_sq_ft	Continuous	Screen porch area in square feet.
pool_area_sq_ft	Continuous	Pool area in square feet.
pool_quality_rating	Ordinal	Pool quality.
fence_quality_rating	Ordinal	Fence quality.
misc_feature	Nominal	Miscellaneous feature not covered in other categories.
misc_value	Continuous	Dollar value of miscellaneous feature.
month_sold	Discrete	Month Sold (MM).
year_sold	Discrete	Year Sold (YYYY).
sale_type	Nominal	Type of sale.
sale_condition	Nominal	Condition of sale.
sale_price	Continuous	Sale price (\$).
latitude	Continuous	Latitude of the property.
longitude	Continuous	Longitude of the property.

See the original <u>data documentation</u> for more details.

Good luck!

Copyright and license

De Cock, D. (2011). <u>Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project</u>, Journal of Statistics Education, Volume 19, Number 3.

Kuhn M (2020). AmesHousing, The Ames Iowa Housing Data. R package version 0.0.4.