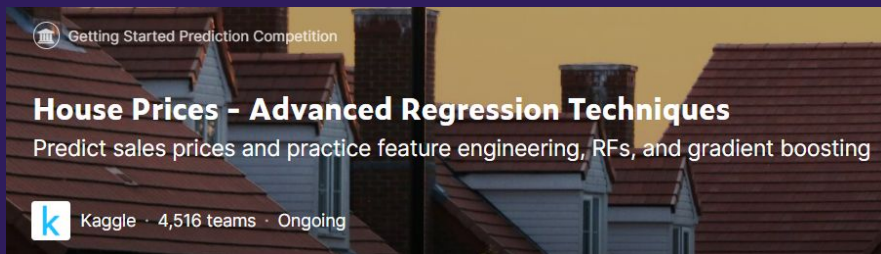




HOUSE PRICES DATA ANALYSIS

2022

Dataset overview



- 1460 samples.
- 80+ features | We moved to 10 for simplicity.
- Target feature: SalePrice.
- Different dtypes.

- 10 chosen features:
- OverallQual: rates overall material and finish of the house.
 - GrLivArea: Above ground living area square feet.
 - GarageCars: Size of garage in car capacity.
 - GarageArea: Size of garage in square feet.
 - TotalBsmtSF: Total square feet of basement area.
 - FullBath: Full bathrooms.
 - YearBuilt: Year built.
 - YearRemodAdd: Remodel date.
 - MSSubClass: Identifies the type of dwelling involved in the sale.
 - MoSold: Month Sold.

BASIC GOALs

- 1) Understand which way many factors can affect certain feature on practice.
- 2) Learn how to analyze data using math knowledge.
- 3) Process dataframe to feed it to ML model.

PLAN

01

CLEAN OUTLIERS

Transform target variable.

03

DATA PREPARATION

Splitting, encoding, scaling etc.

02

FEATURE ENGINEERING

Decide if features affect target.

04

FEATURE IMPORTANCE VERIFICATION & ML

Random forest and boosting.