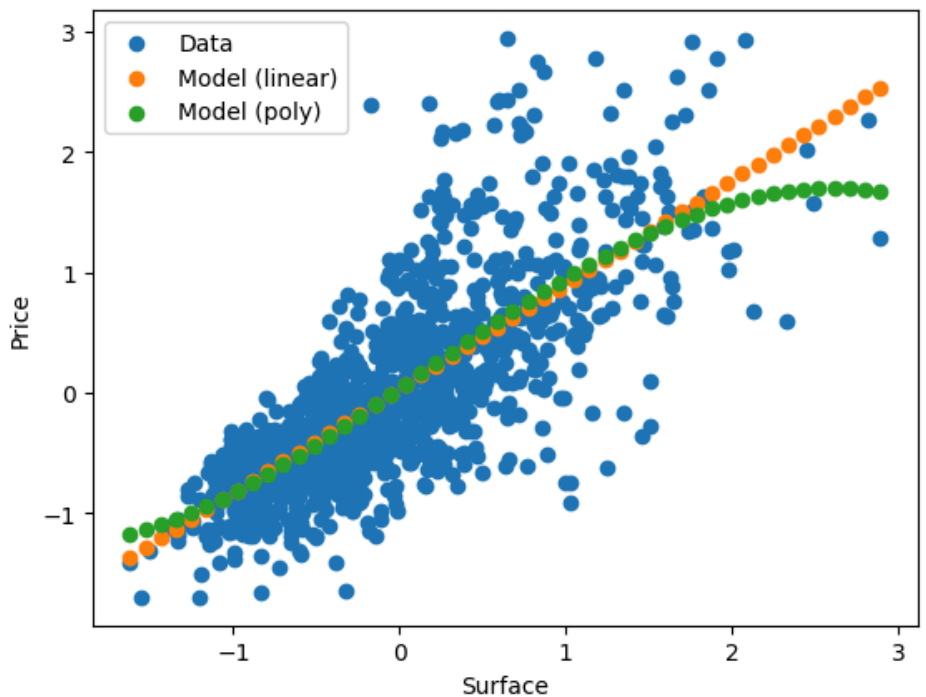
Business Question

What is the value of a house, given its different characteristics ?



Dataset Presentation

Preliminary analysis and possible models

Public data on house sales in a USA county

Gives an idea of what the housing market looks like

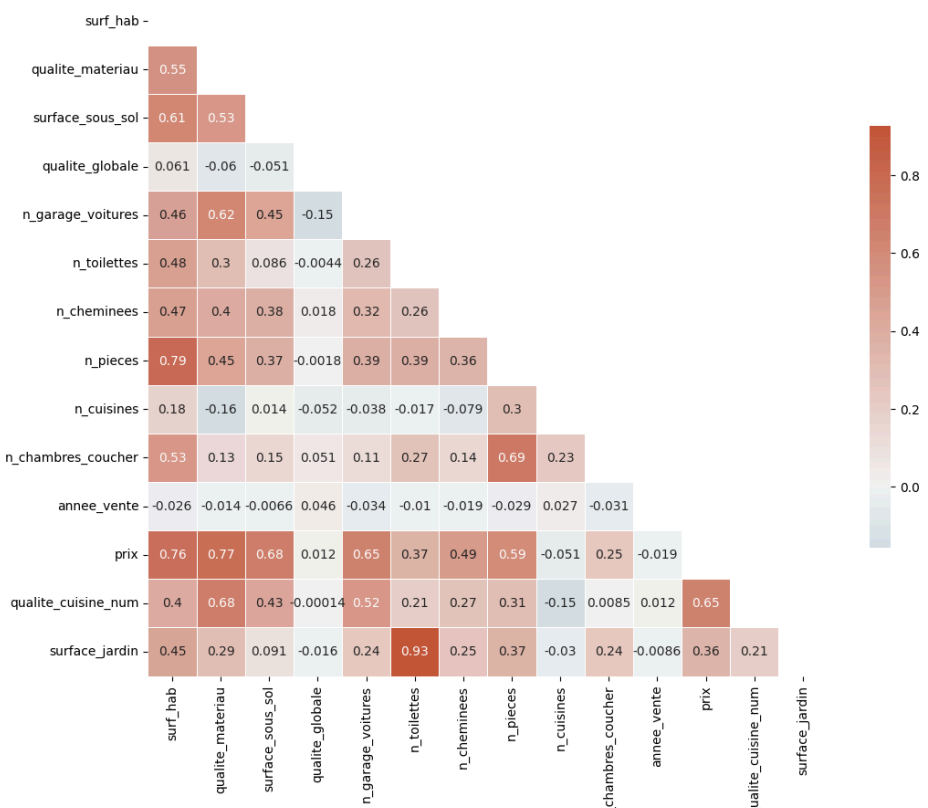
1364 entrypoints, each with 15 characteristics

Data Description

- (N) Surface
- (N) Materials quality
- (N) Underground surface
- (N) Global quality
- (N) # of parking spaces
- (N) # of toilets
- (N) # of fireplaces
- (N) # of rooms
- (N) # of kitchens
- (N) # of bedrooms
- (N) Sale year
- (N) Price
- (C) Type of roof
- (C) Type of house
- (N) Kitchen quality
- (N) Garden surface

(N) = Numerical
(C) = Categorical

Exploratory Data Analysis



High correlations with price

- materials quality (0.77)
- surface (0.76)
- underground surface (0.68)
- number of parking spaces (0.65)
- number of rooms (0.59)

Correlation between garden surface and # of toilets: 0.93!

Modelling

Polynomial model of degree 6 based on multiple variables:

- Surface
- Materials quality
- # of rooms
- # of parking spaces
- “Is premium” (whether global quality > 6/10)

Model Performance and Validatation

	Linear Model	Polynomial Model
Training Score[1]	CHF 37′861.152	CHF 26′027.110
Validation Score[1]	CHF 37′032.211	CHF 24′633.376

[1] Mean absolute error

Business Conclusions and Next Steps

TODO