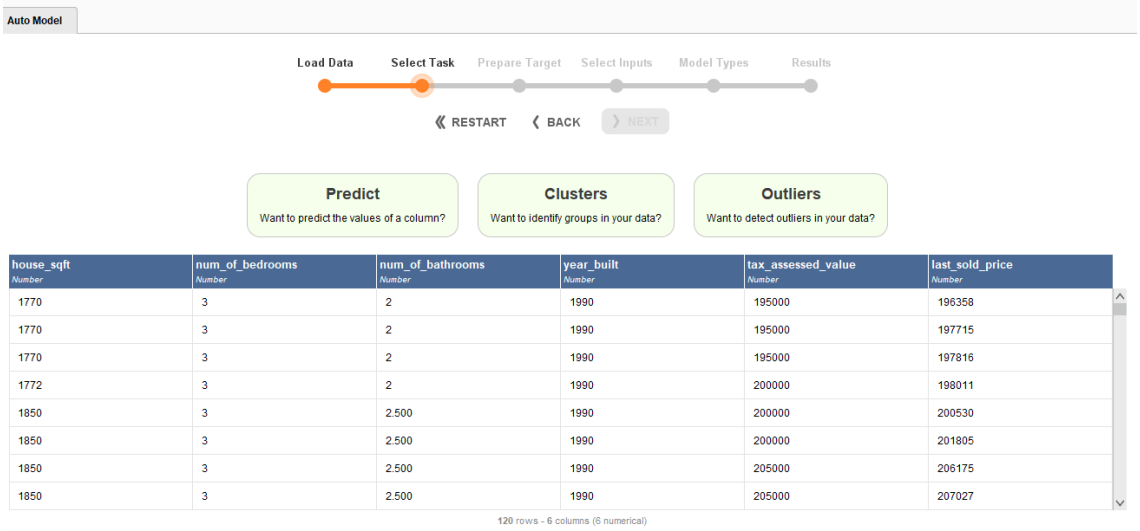
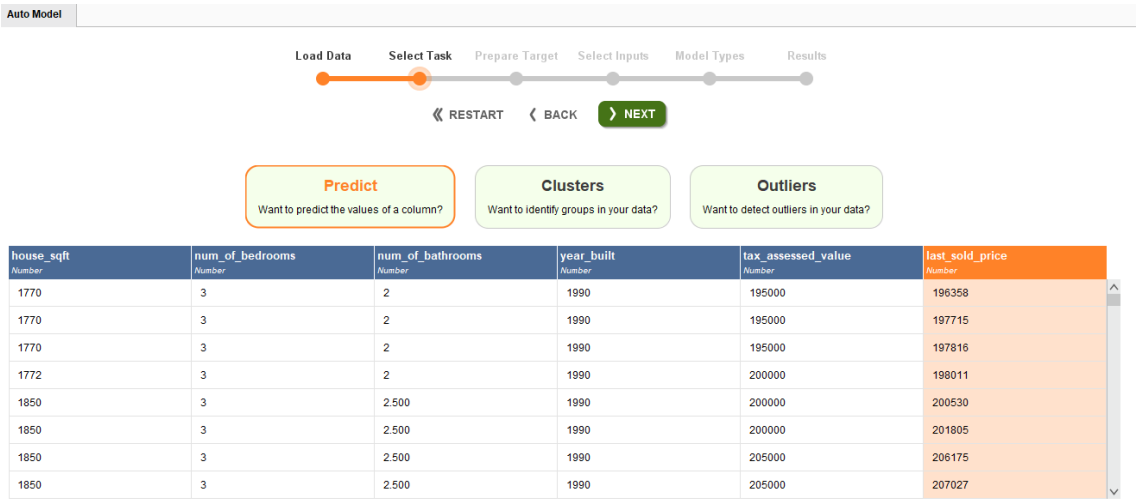
ANALÍTICA DE DATOS

Estudiante: Danny Sebastián Díaz Padilla

Taller

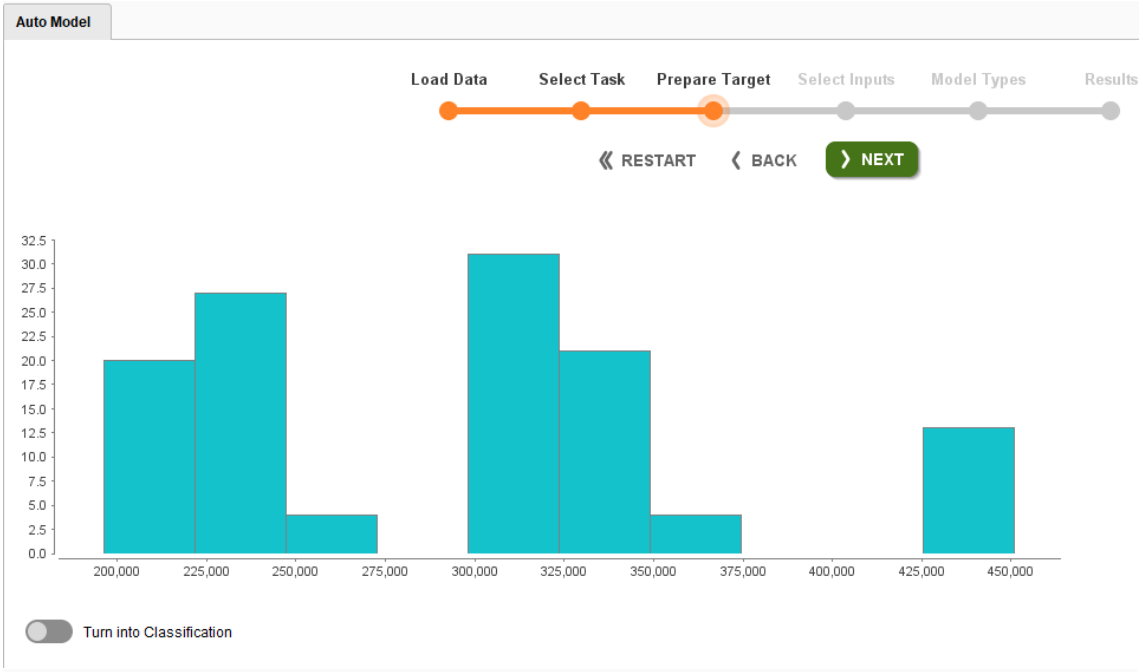


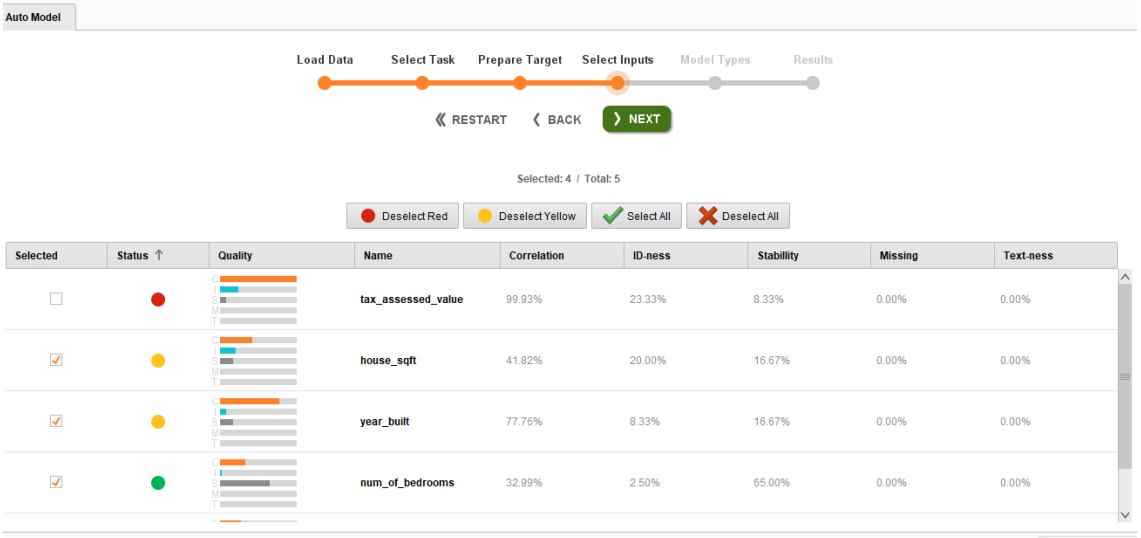
house_sqft Number	num_of_bedrooms Number	num_of_bathrooms Number	year_built Number	tax_assessed_value Number	last_sold_price Number
1770	3	2	1990	195000	196358
1770	3	2	1990	195000	197715
1770	3	2	1990	195000	197816
1772	3	2	1990	200000	198011
1850	3	2.500	1990	200000	200530
1850	3	2.500	1990	200000	201805
1850	3	2.500	1990	205000	206175
1850	3	2.500	1990	205000	207027

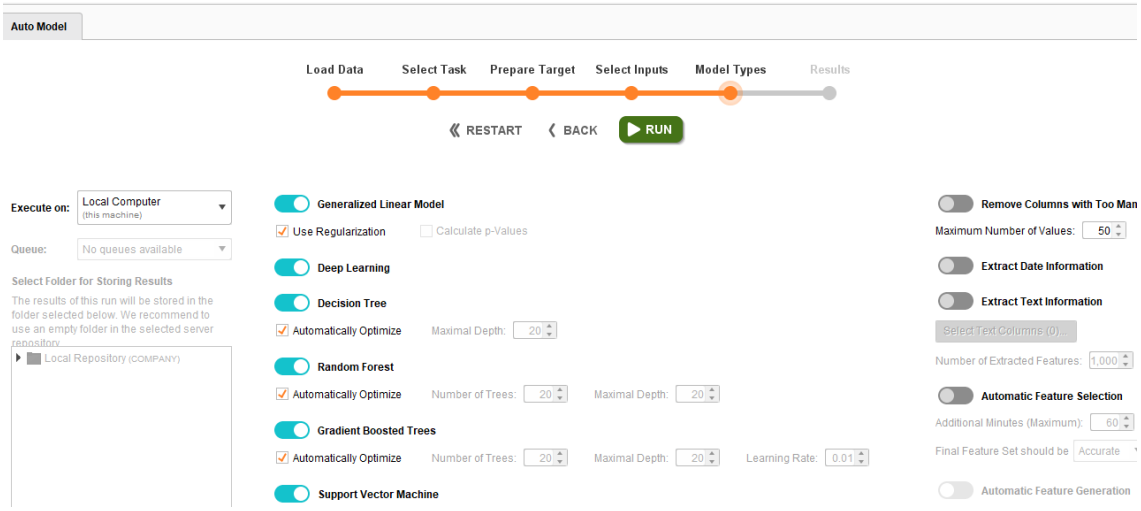


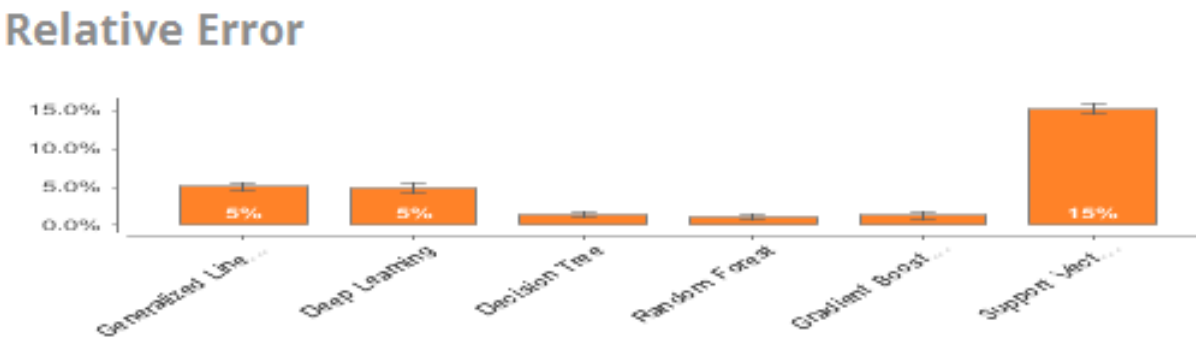
house_sqft Number	num_of_bedrooms Number	num_of_bathrooms Number	year_built Number	tax_assessed_value Number	last_sold_price Number
1770	3	2	1990	195000	196358
1770	3	2	1990	195000	197715
1770	3	2	1990	195000	197816
1772	3	2	1990	200000	198011
1850	3	2.500	1990	200000	200530
1850	3	2.500	1990	200000	201805
1850	3	2.500	1990	205000	206175
1850	3	2.500	1990	205000	207027

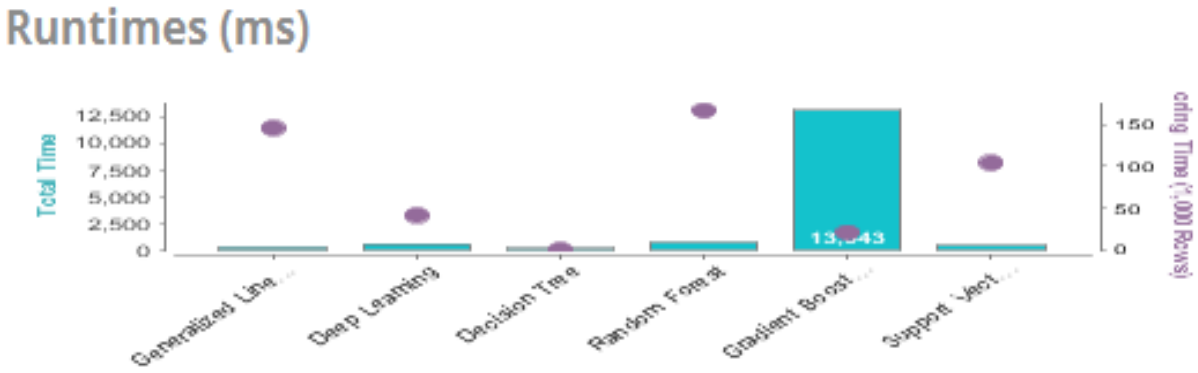
ANALÍTICA DE DATOS











Model	Relative Error	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)
Generalized Linear Model	5.1%	± 0.4%	?	254 ms	425 ms
Deep Learning	4.9%	± 0.7%	?	596 ms	2 s
Decision Tree	1.4%	± 0.3%	?	136 ms	8 ms

Model	Relative Error	Standard Deviation	Gains	Total Time	Training Time (1,000 Rows)
Random Forest	1.1%	± 0.2%	?	697 ms	100 ms
Gradient Boosted Trees	1.3%	± 0.5%	?	13 s	3 s
Support Vector Machine	15.1%	± 0.6%	?	406 ms	92 ms

Importar data de casas en venta de una Inmobiliaria

house_sqft – superficie de la casa
num_of_bedrooms - Número de habitaciones
num_of_bathrooms - Número de baños
year_built - El año en que se construyó la casa
tax_assessed_value - Valor de acceso fiscal de la casa
last_sold_price - Último precio de venta de la casa
rate_per_sqfoot – Tasa por pie cuadrado.
home_type – (value apartment, townhouse or single family home)
school_rating_1to10 – Calificación de la escuela que le corresponde al lugar

ANALÍTICA DE DATOS

Identificar los algoritmos con mejor resultados

El modelo con mayor velocidad al otorgar un score y en general es el Árbol de decisiones con solo un 1.4% de error.

Decision Tree	1.4%	± 0.3%	?	136 ms	8 ms

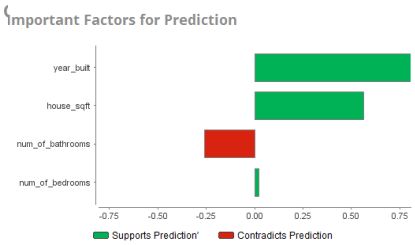
El mejor modelo en cuestión de predicción es “Random Forest” con una menor desviación que el árbol de decisiones pero se demora 5.125 (o mejor dicho 697/136) veces más al entrenar y 12.5 (o mejor dicho 100/8) veces más al otorgar scores.

Random Forest	1.1%	± 0.2%	?	697 ms	100 ms

Predecir 10 distintos escenarios de venta de casas con distintos algoritmos

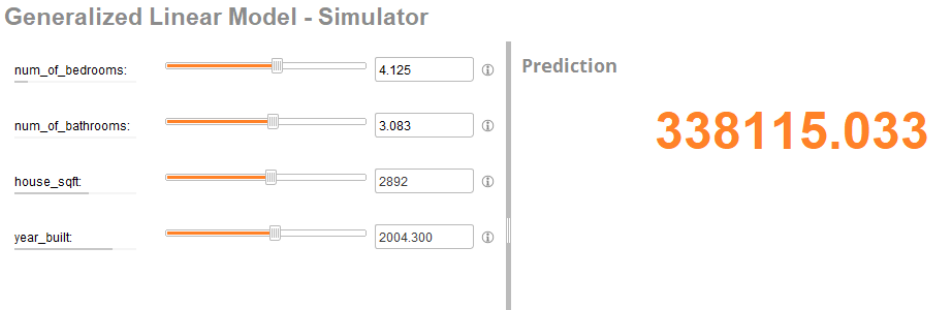
A. Algoritmo: Generalized Linear Model

Factores:

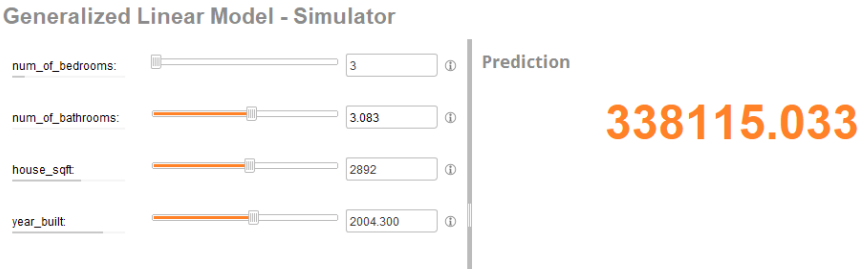


Escenarios:

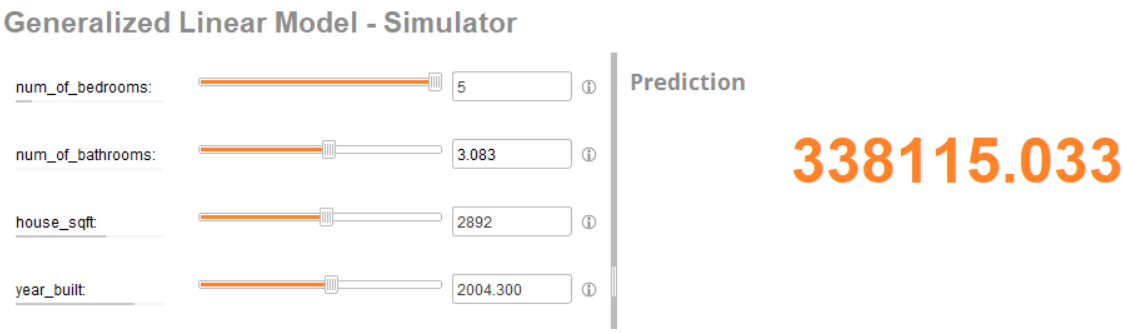
1. Con número promedio de datos.



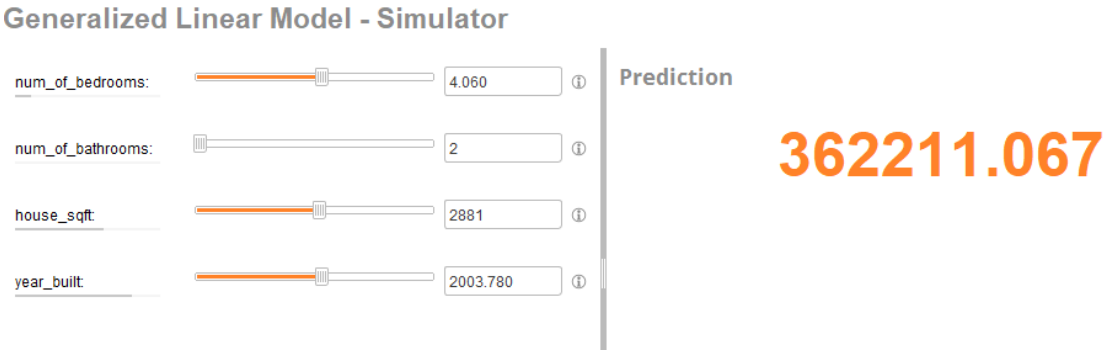
2. Con bajo número de cuartos



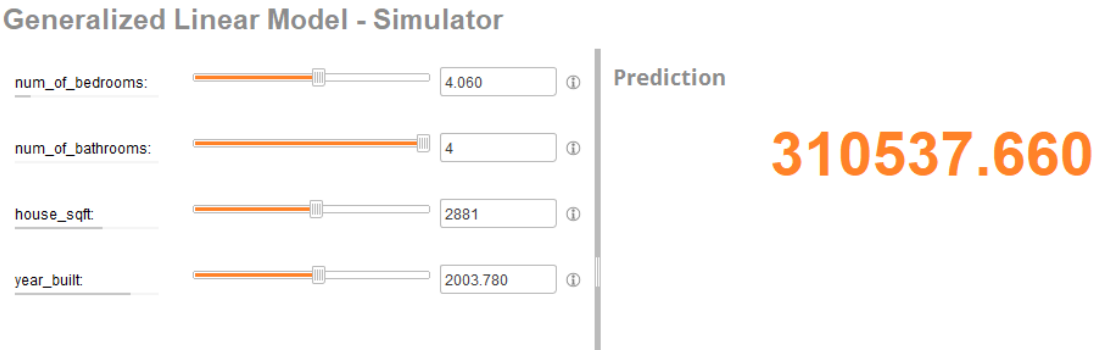
3. Con alto número de cuartos



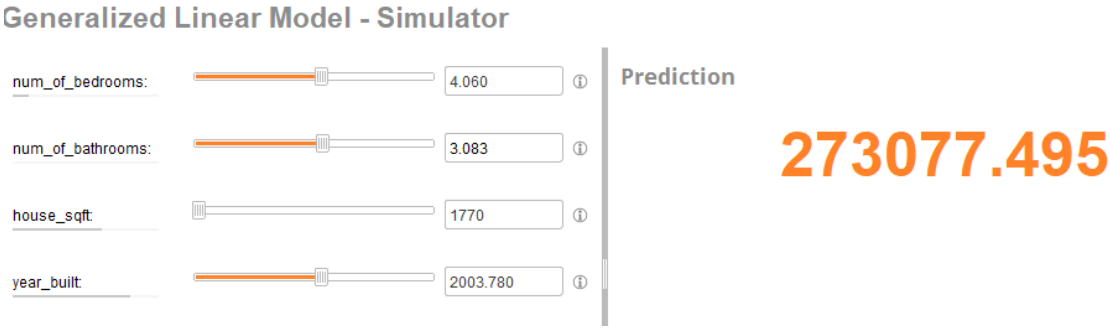
4. Con bajo número de baños



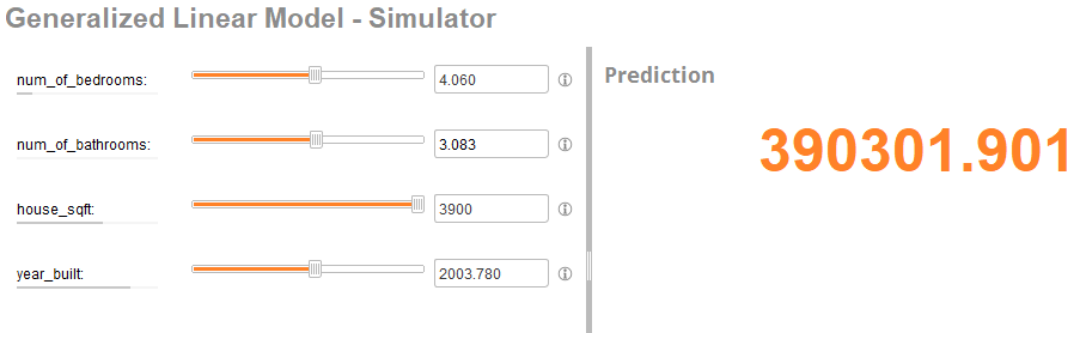
5. Con alto número de baños



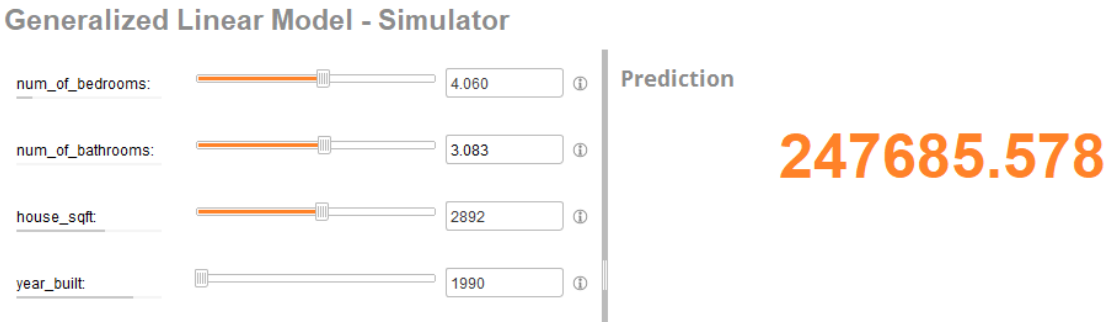
6. Con una superficie pequeña



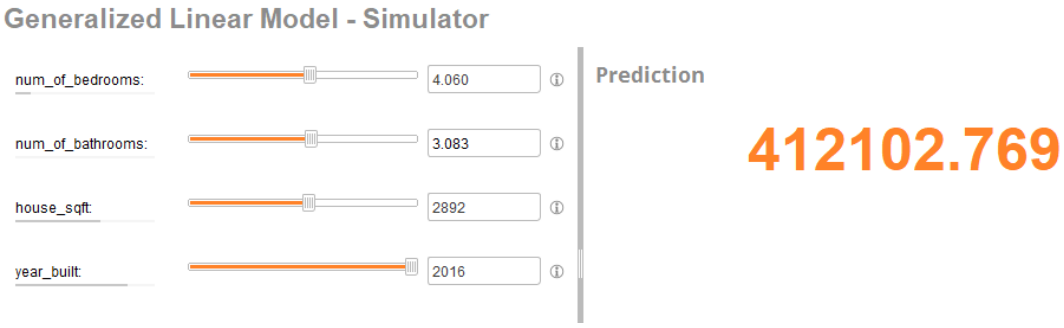
7. Con una superficie enorme



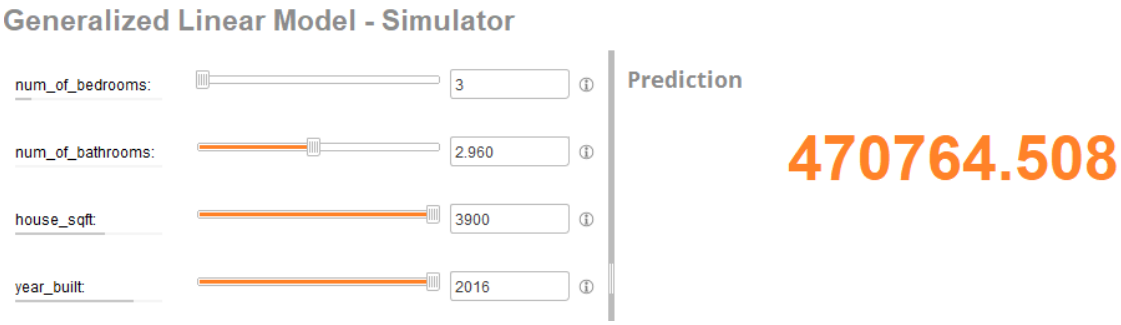
8. Casa antigua



9. Casa moderna

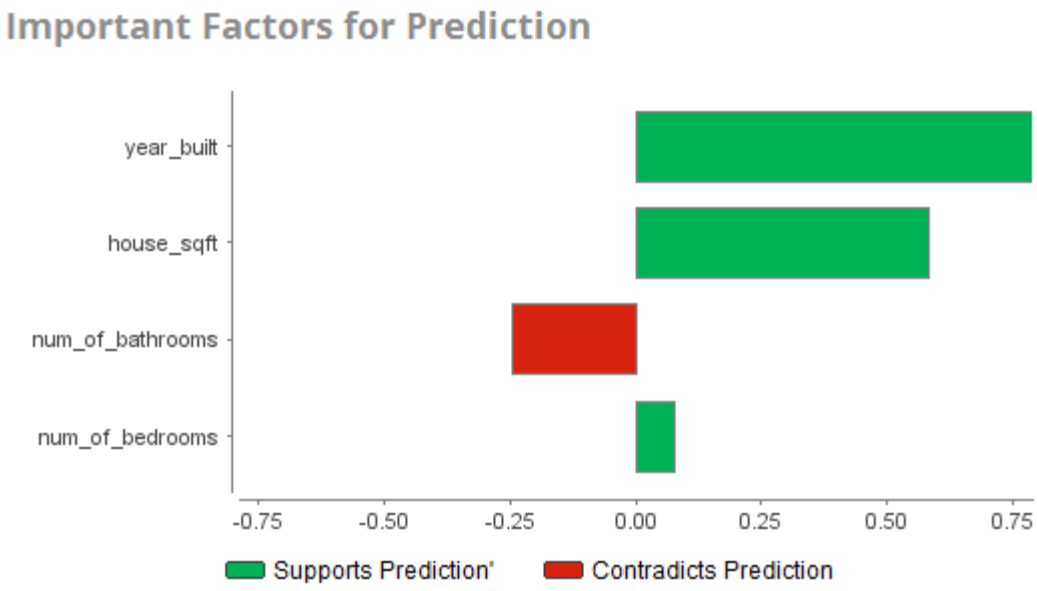


10. Con alta superficie, bajo número de cuartos pero construida recientemente



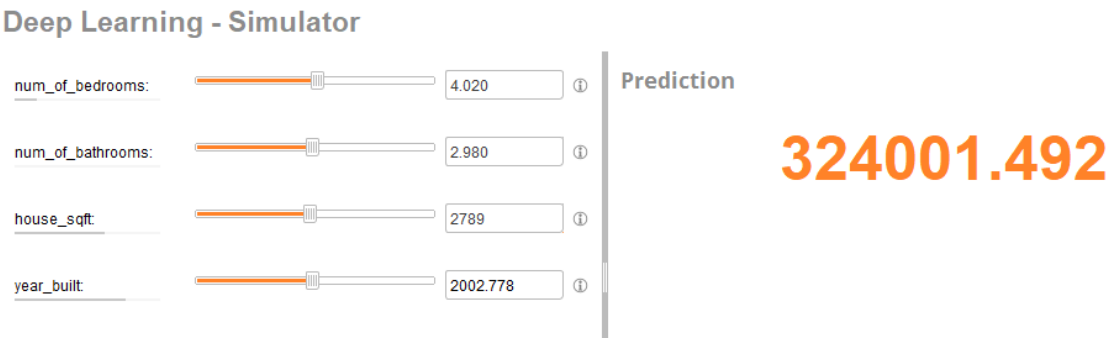
B. Algoritmo: Deep Learning

Factores:

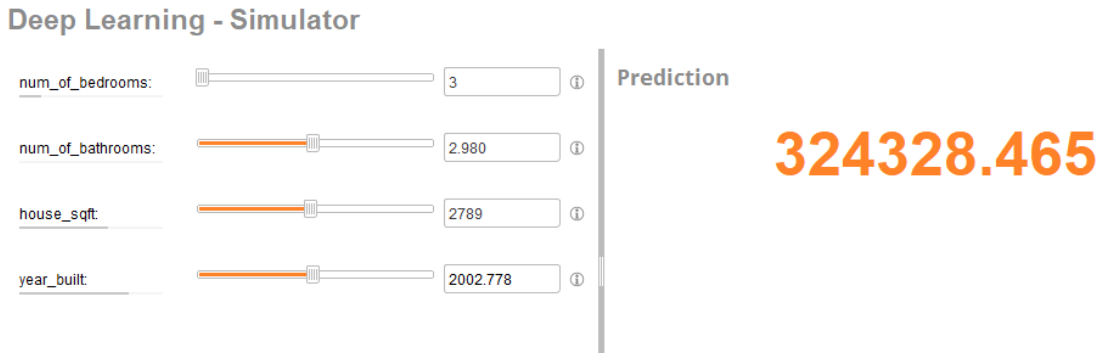


Escenarios:

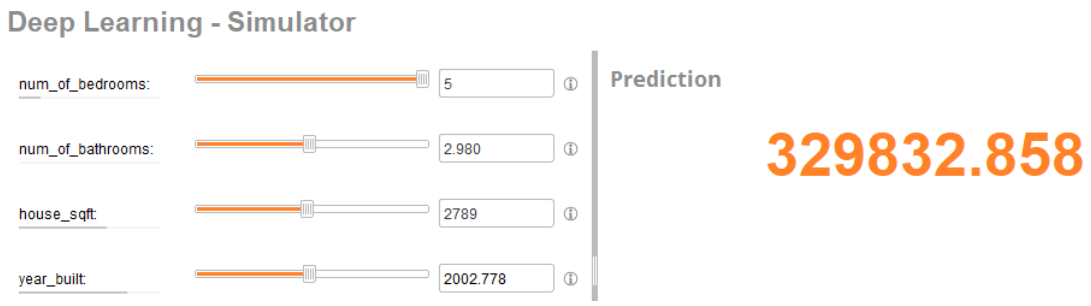
1. Con número promedio de datos



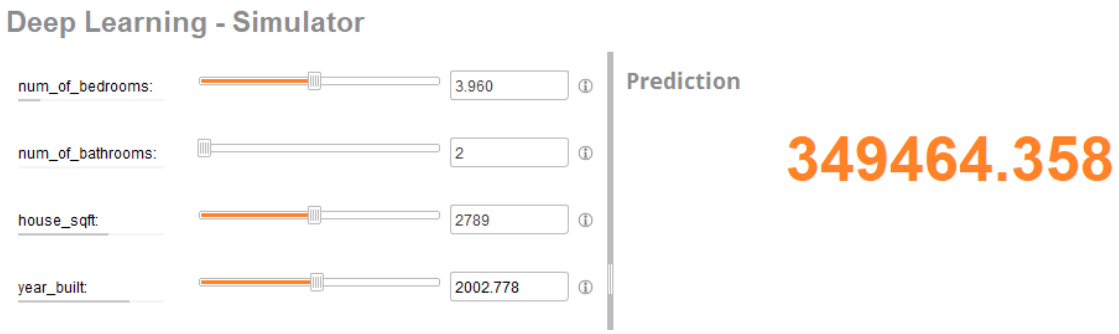
2. Con bajo número de cuartos



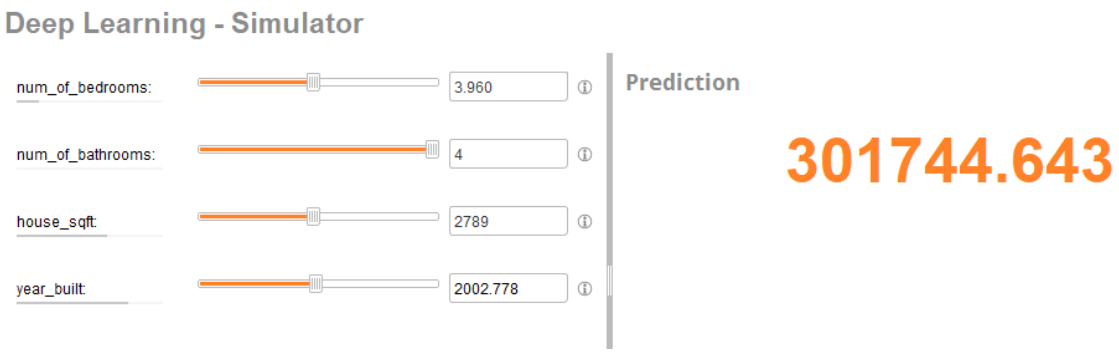
3. Con alto número de cuartos



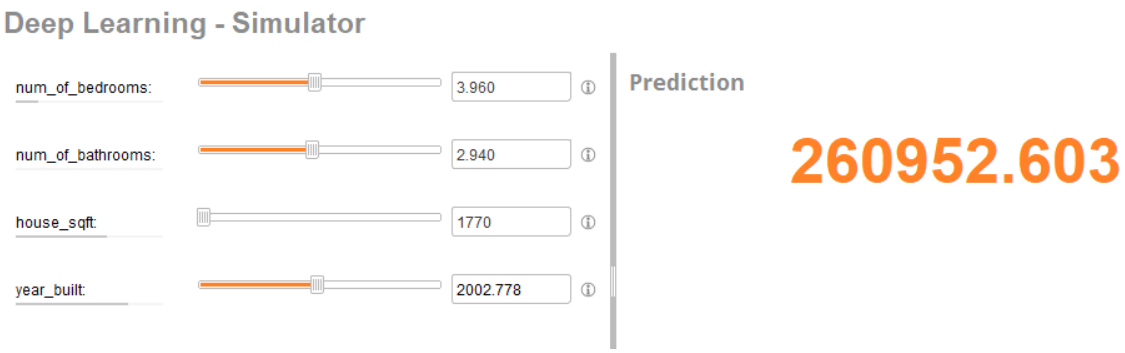
4. Con bajo número de baños



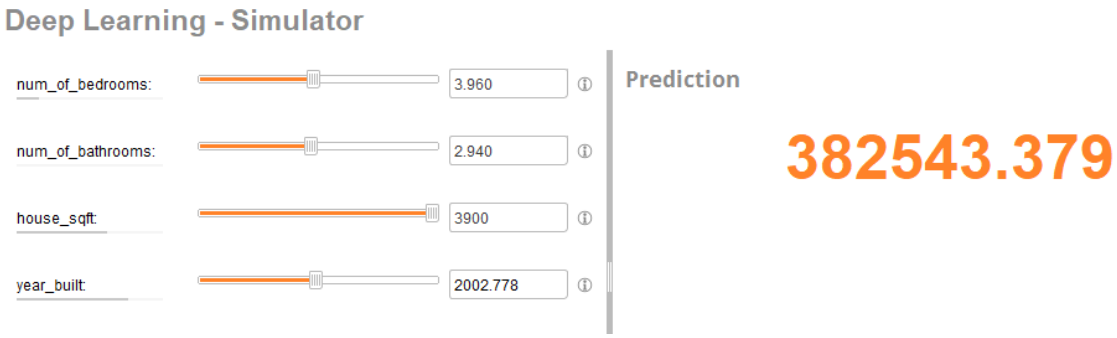
5. Con alto número de baños



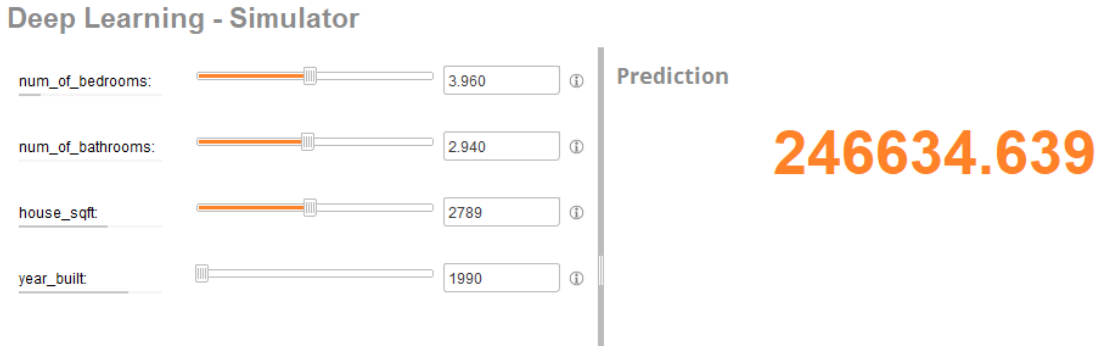
6. Con poca superficie



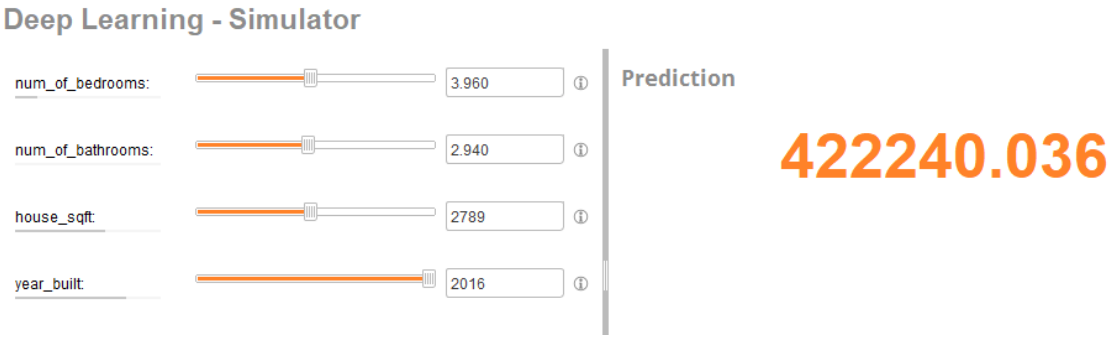
7. Con gran superficie



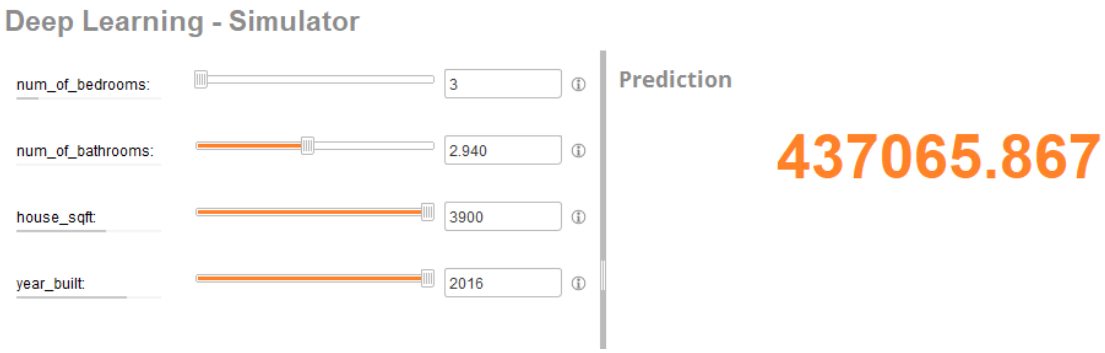
8. Casa antigua



9. Casa moderna

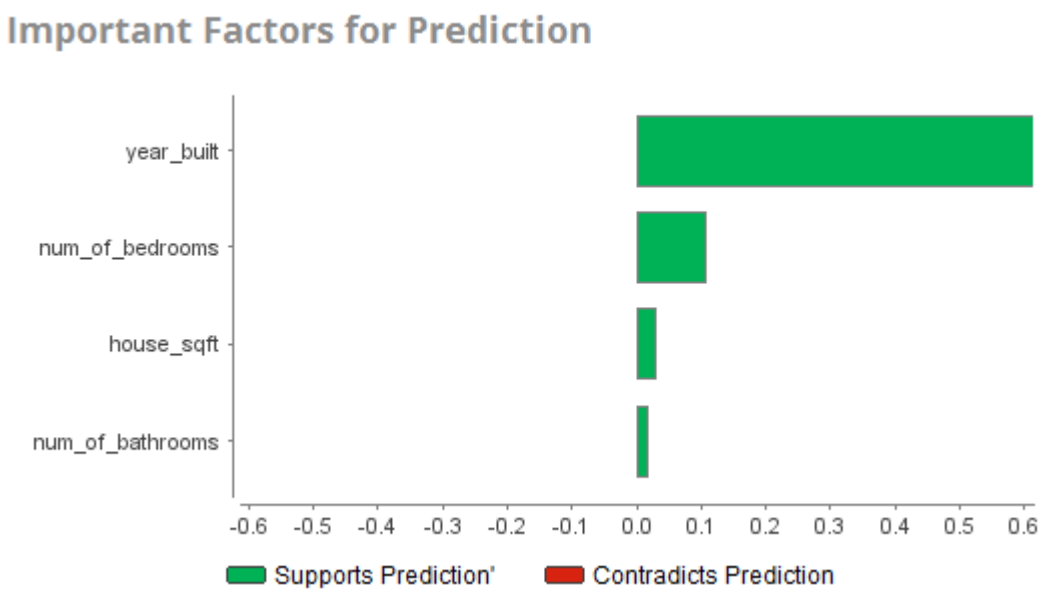


10. Casa con pocos cuartos, construida actualmente y con gran superficie



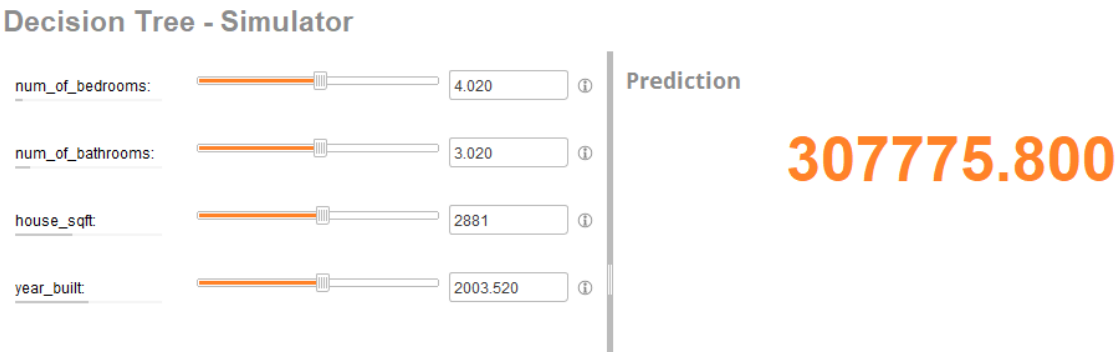
C. Algoritmo: Decision Tree

Factores:

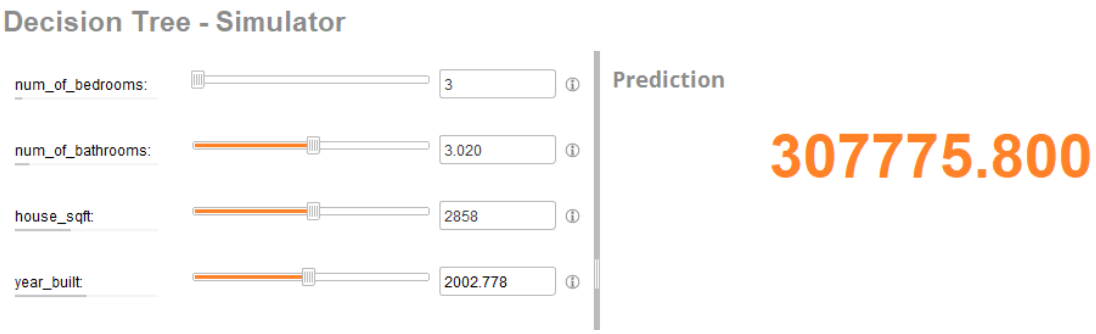


Escenarios:

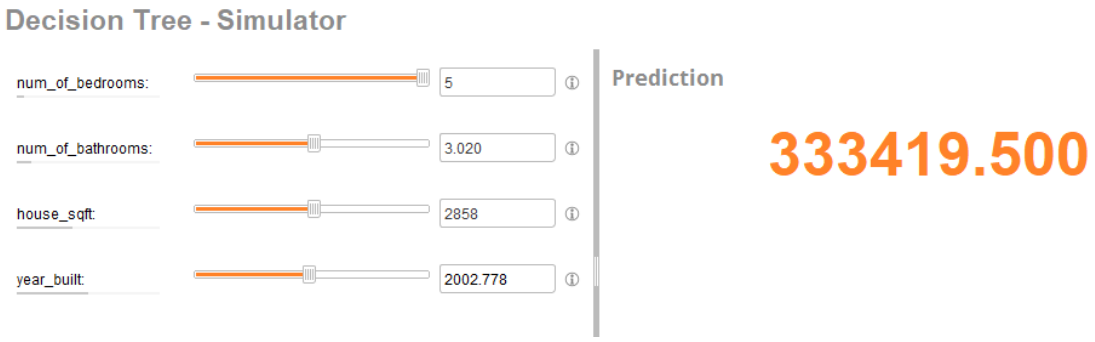
1. Con número promedio de datos



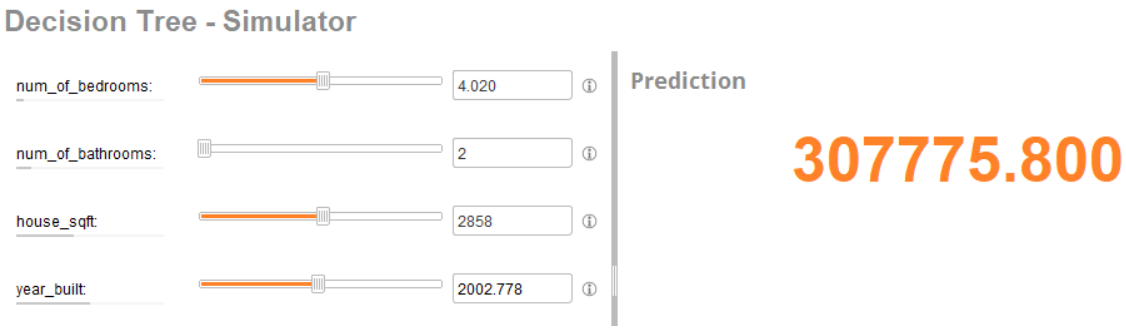
2. Con bajo número de cuartos



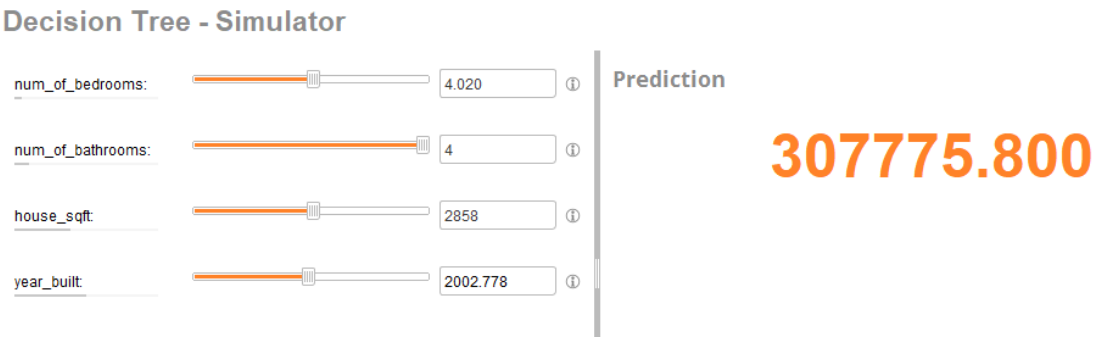
3. Con alto número de cuartos



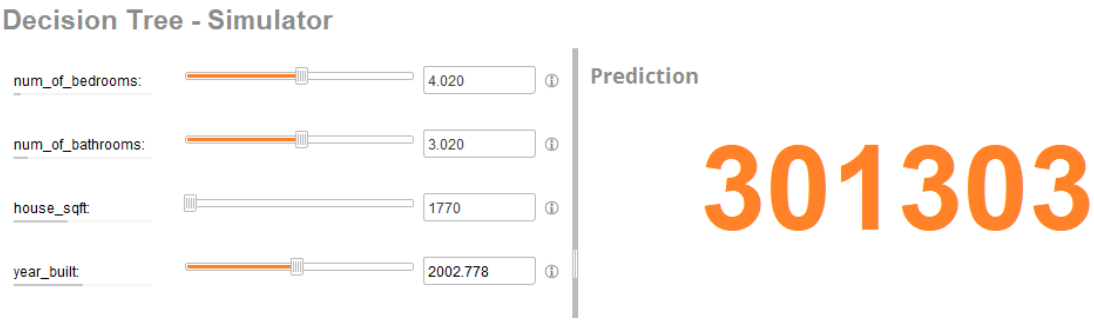
4. Con bajo número de baños



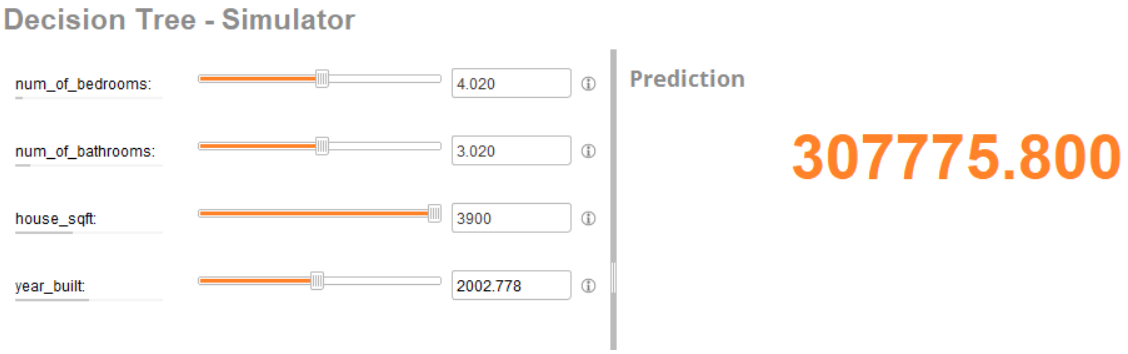
5. Con alto número de baños



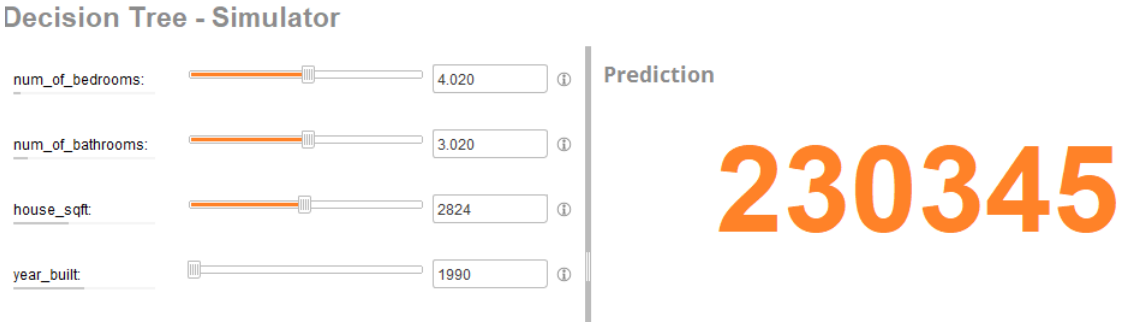
6. Con poca superficie



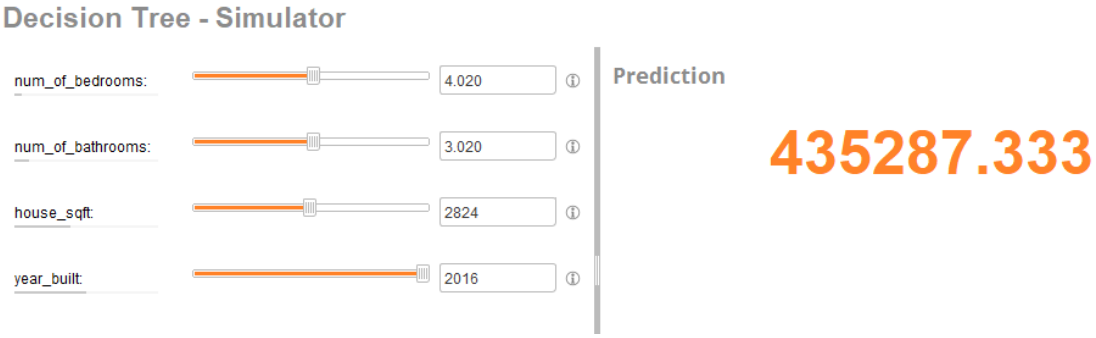
7. Con gran superficie



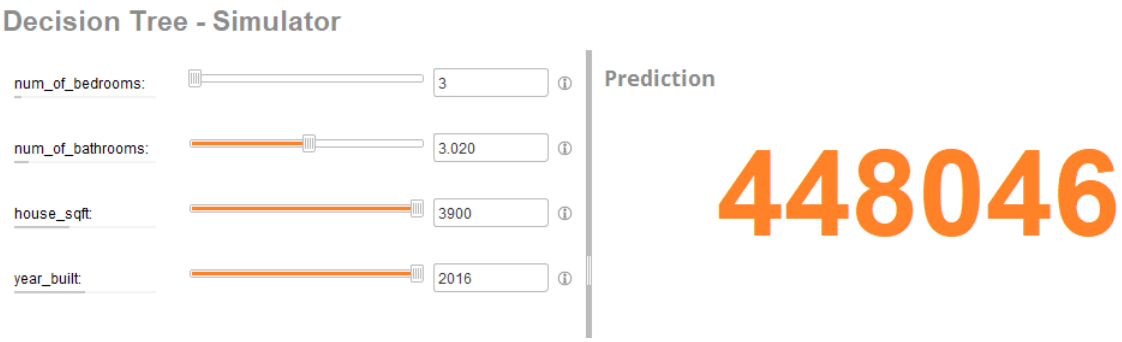
8. Casa antigua



9. Casa moderna

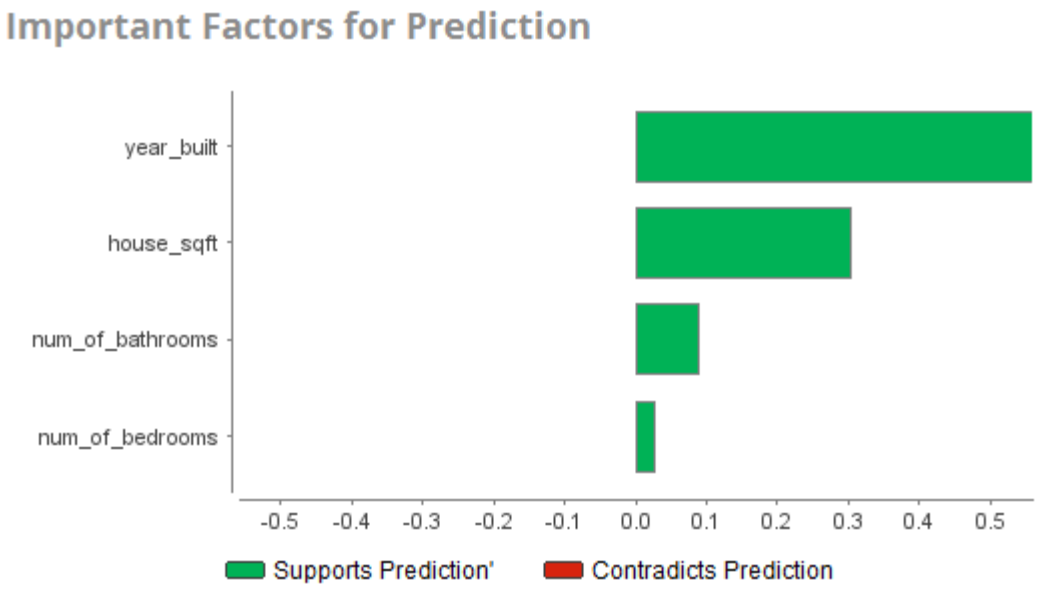


10. Casa con pocos cuartos, construida actualmente y con gran superficie



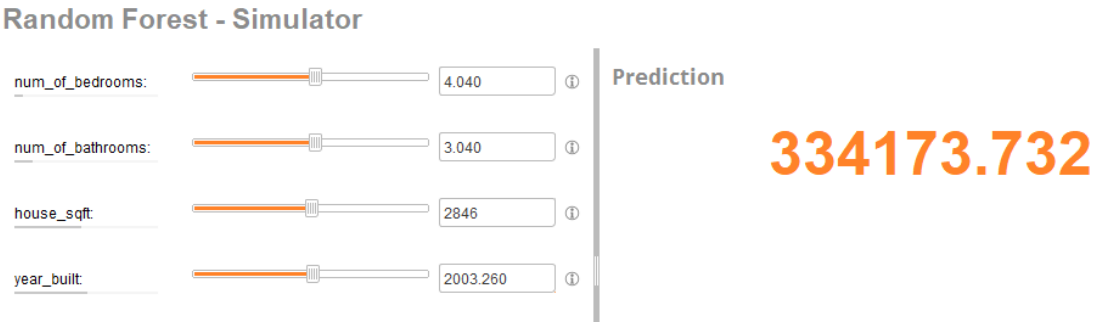
D. Algoritmo: Random Forest

Factores:

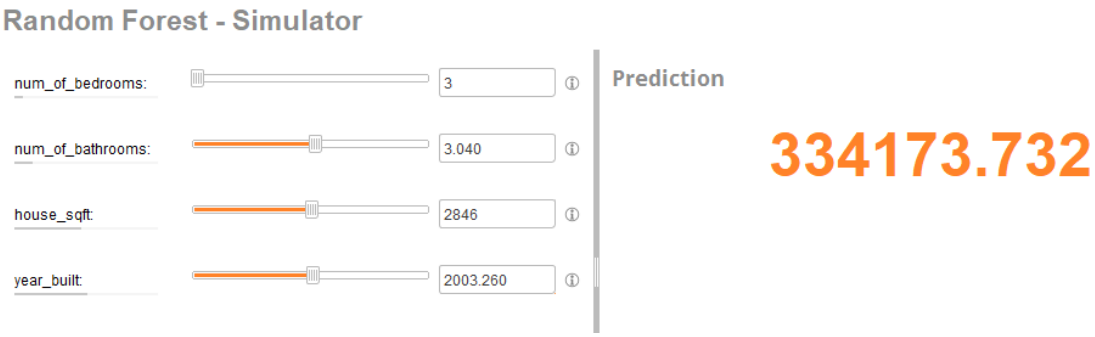


Escenarios:

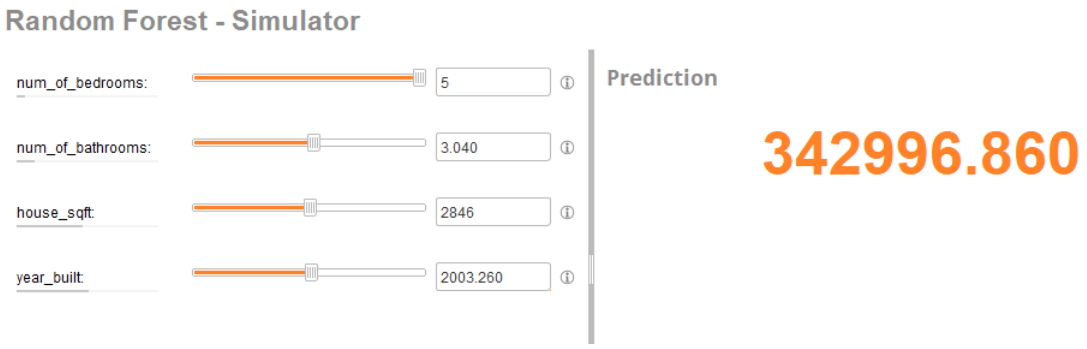
1. Con número promedio de datos



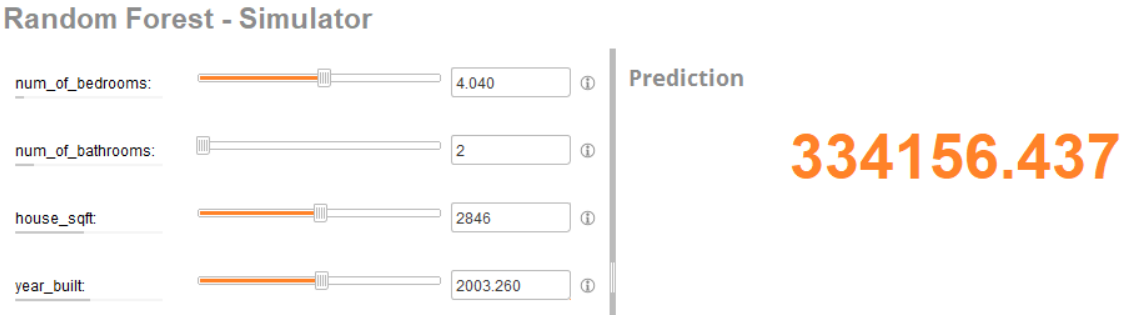
2. Con bajo número de cuartos



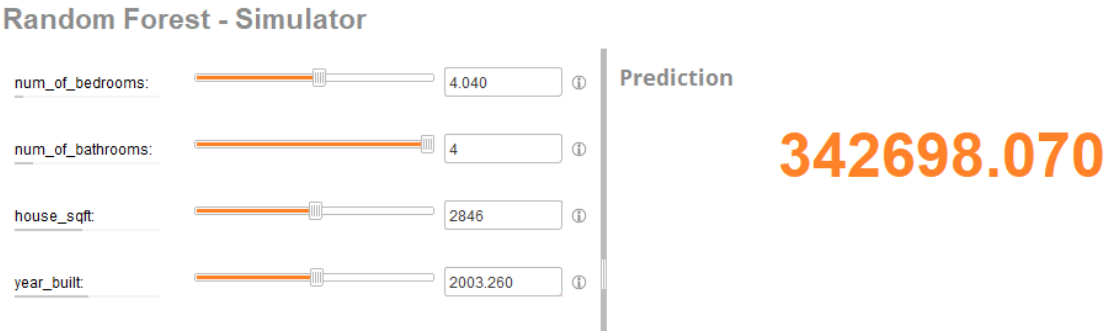
3. Con alto número de cuartos



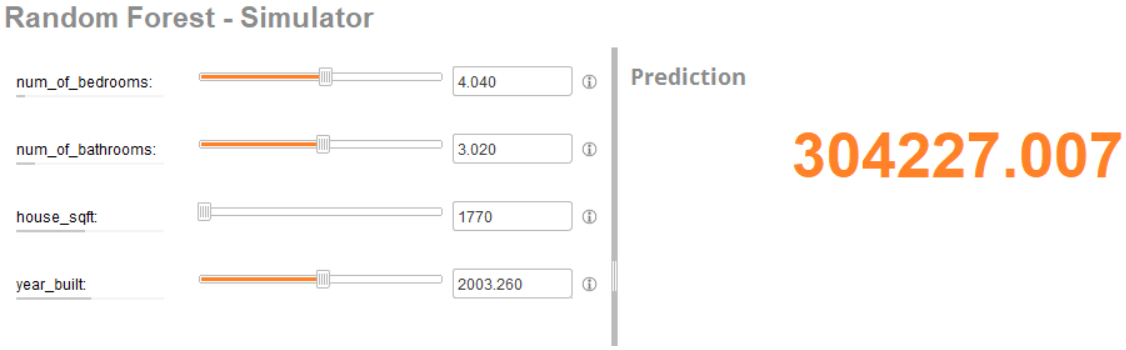
4. Con bajo número de baños



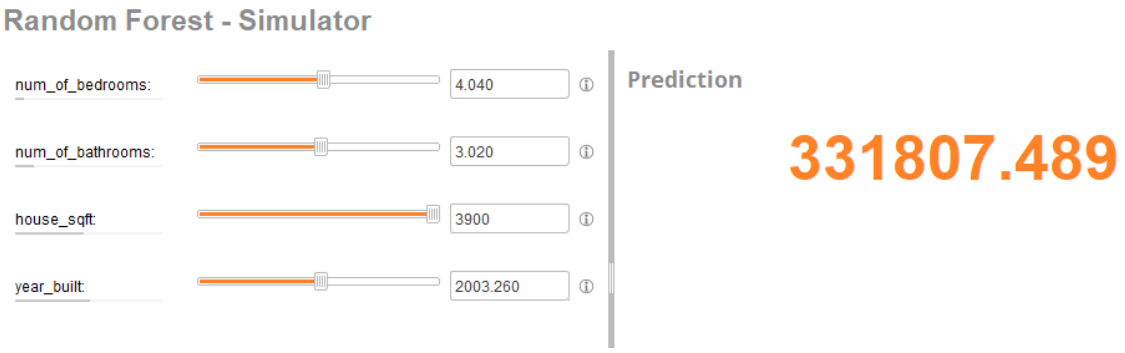
5. Con alto número de baños



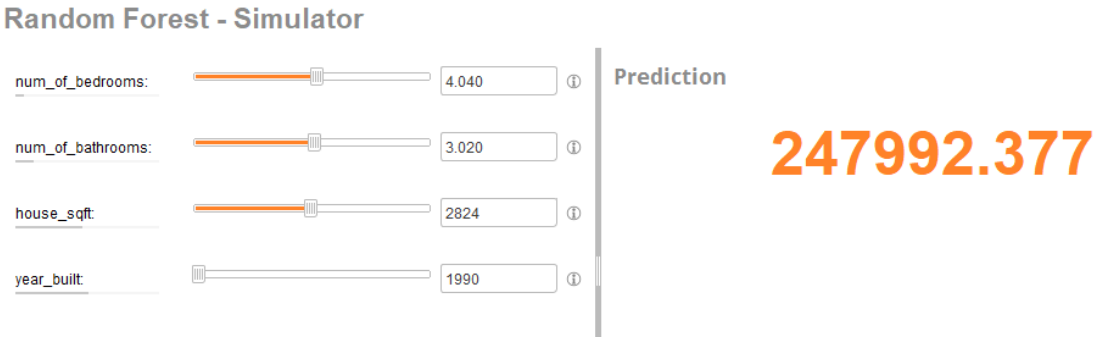
6. Con poca superficie



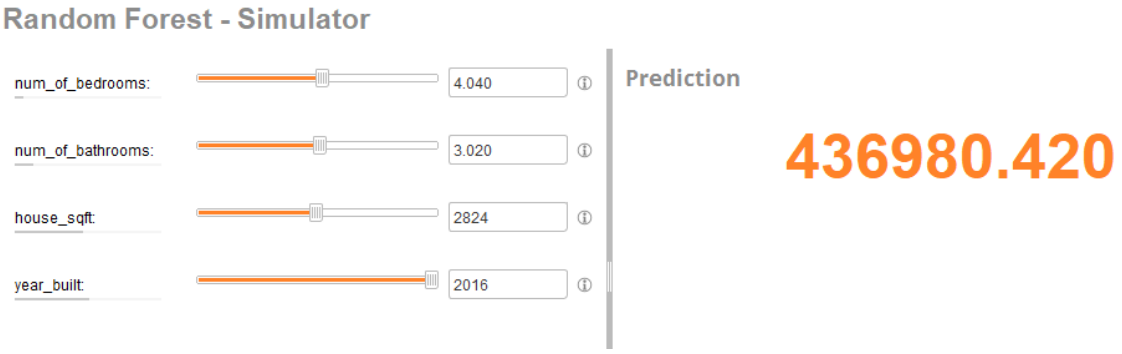
7. Con gran superficie



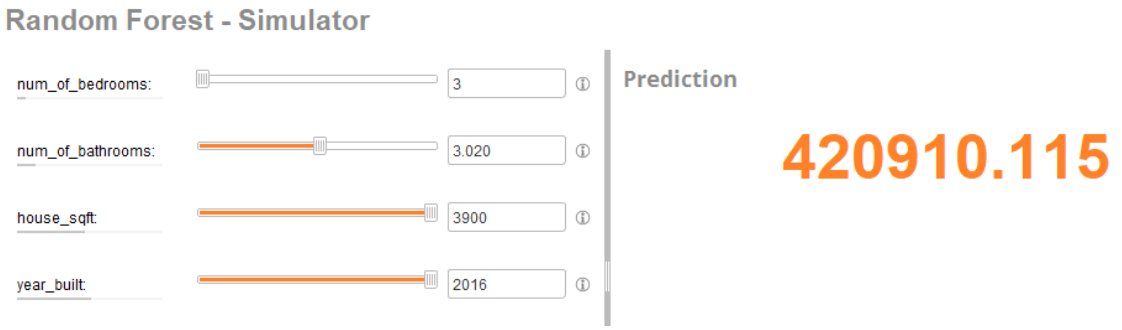
8. Casa antigua



9. Casa moderna



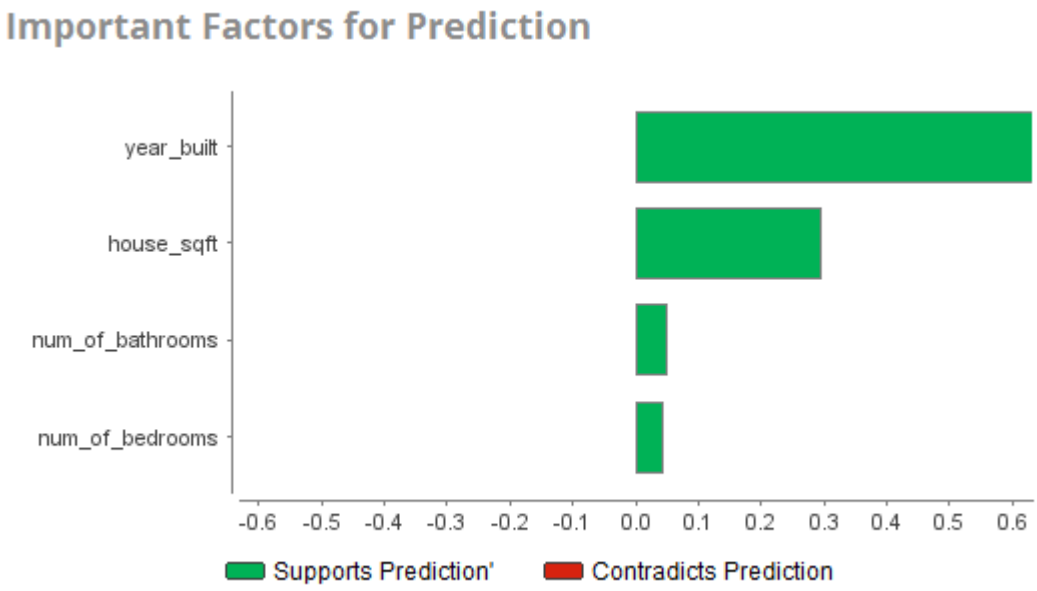
10. Casa con pocos cuartos, construida actualmente y con gran superficie



ANALÍTICA DE DATOS

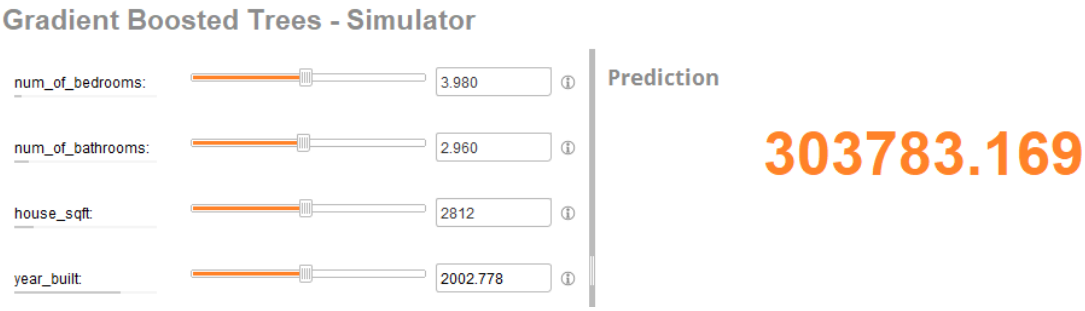
E. Algoritmo: Gradient Boosted Trees

Factores:

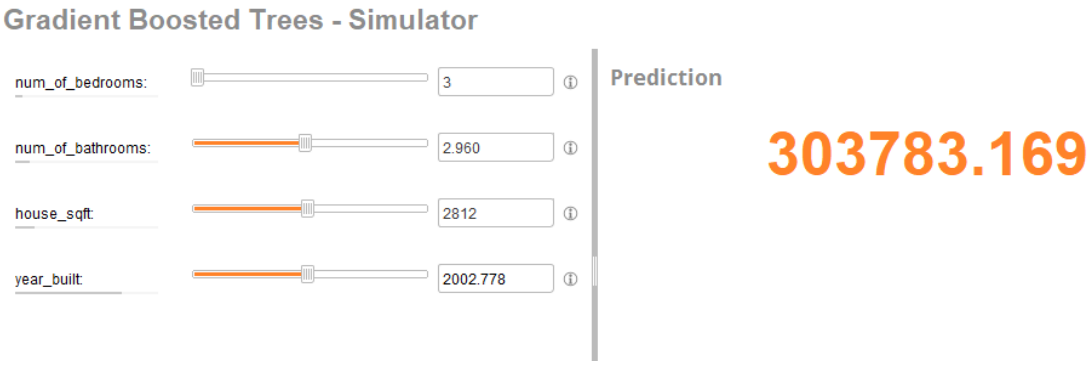


Escenarios:

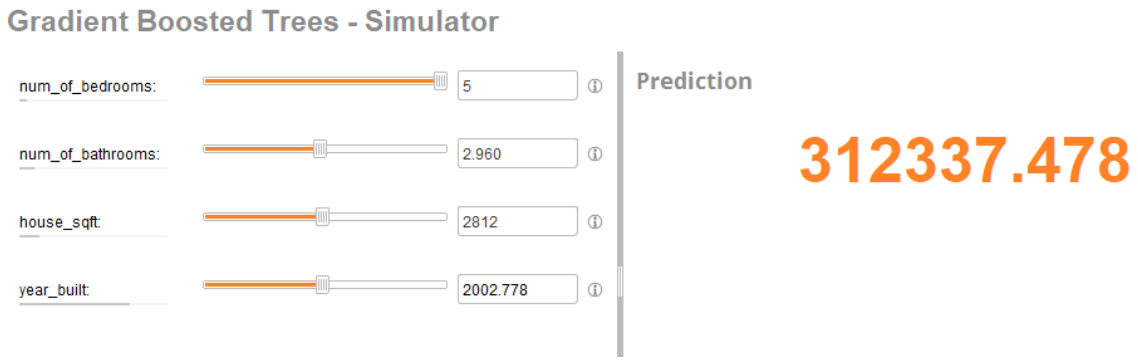
1. Con número promedio de datos



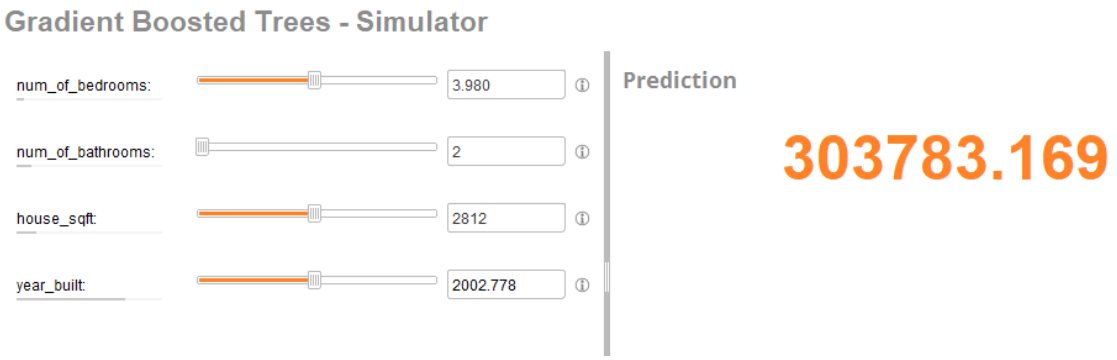
2. Con bajo número de cuartos



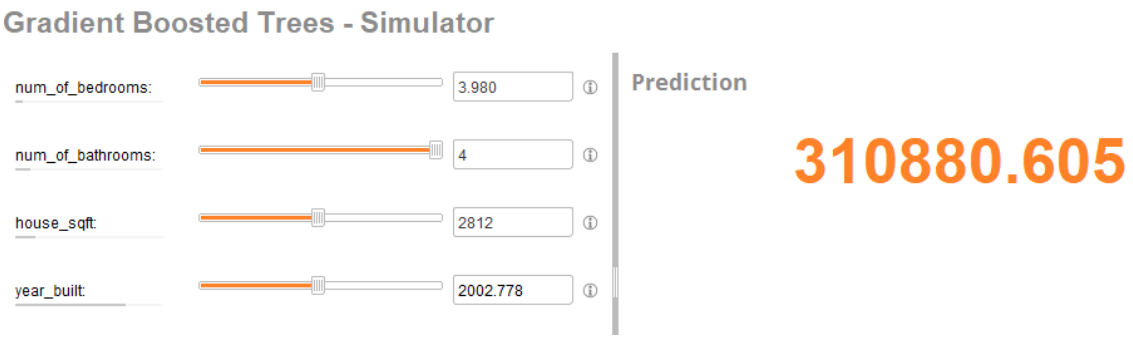
3. Con alto número de cuartos



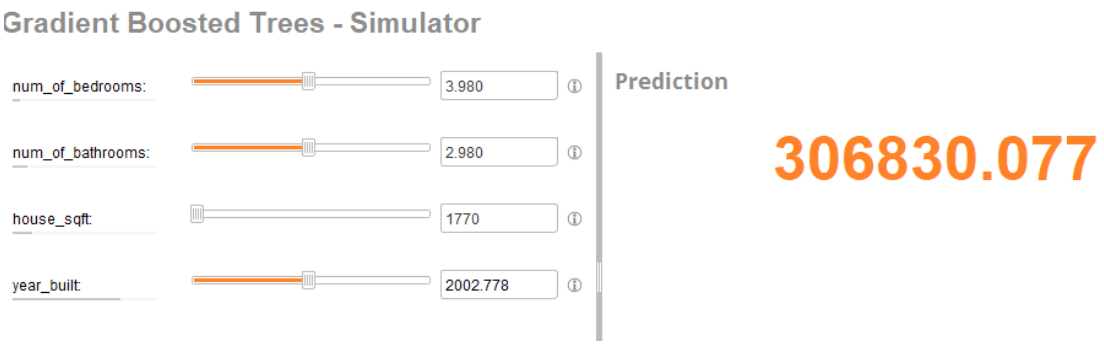
4. Con bajo número de baños



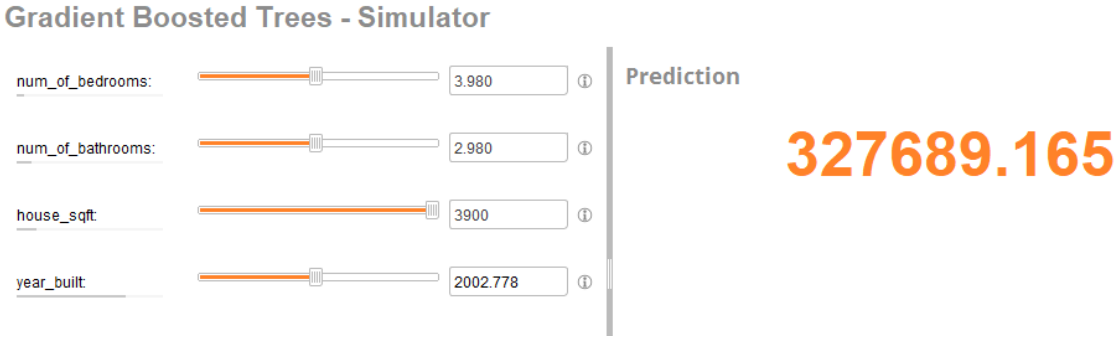
5. Con alto número de baños



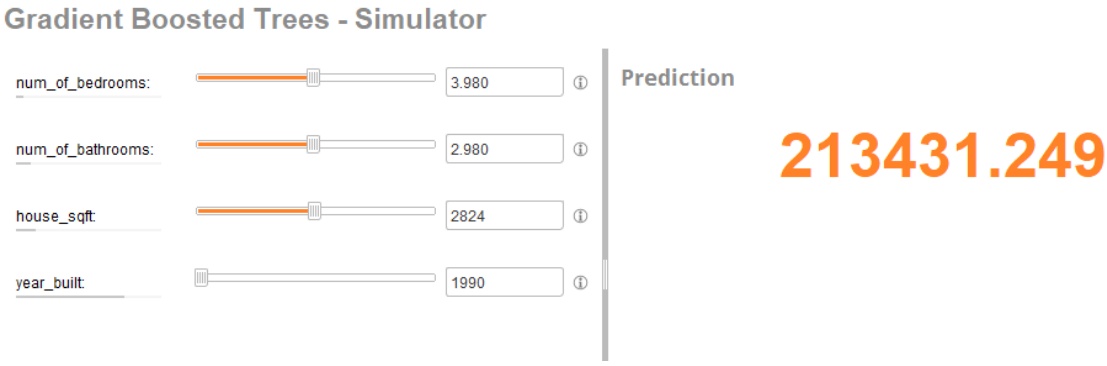
6. Con poca superficie



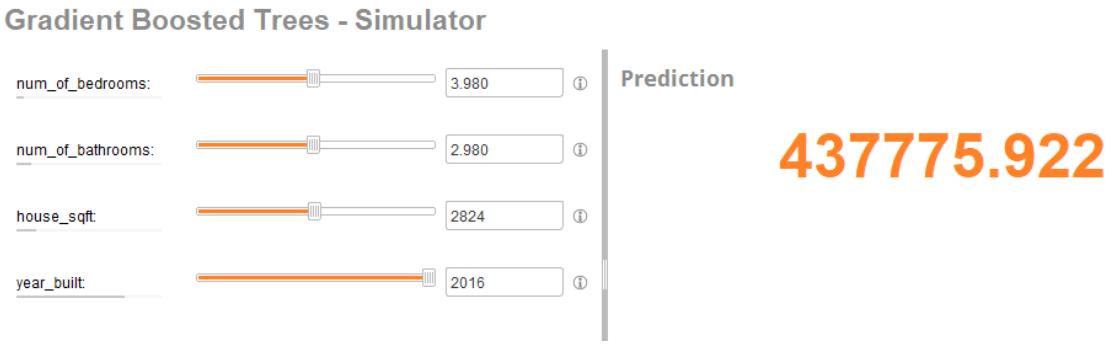
7. Con gran superficie



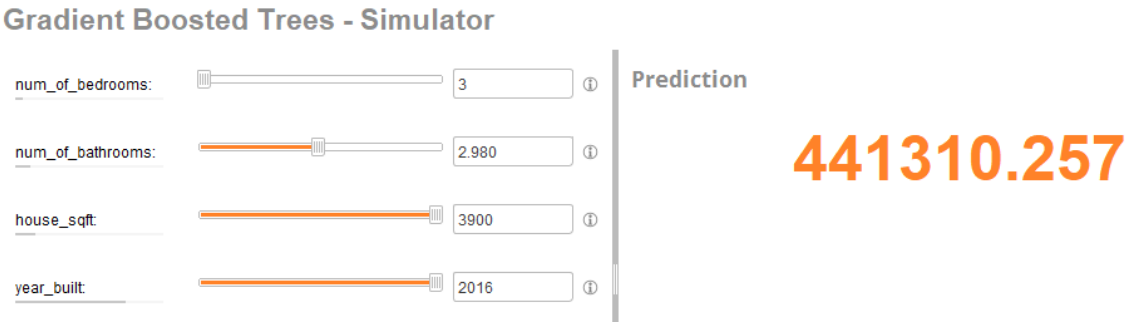
8. Casa antigua



9. Casa moderna



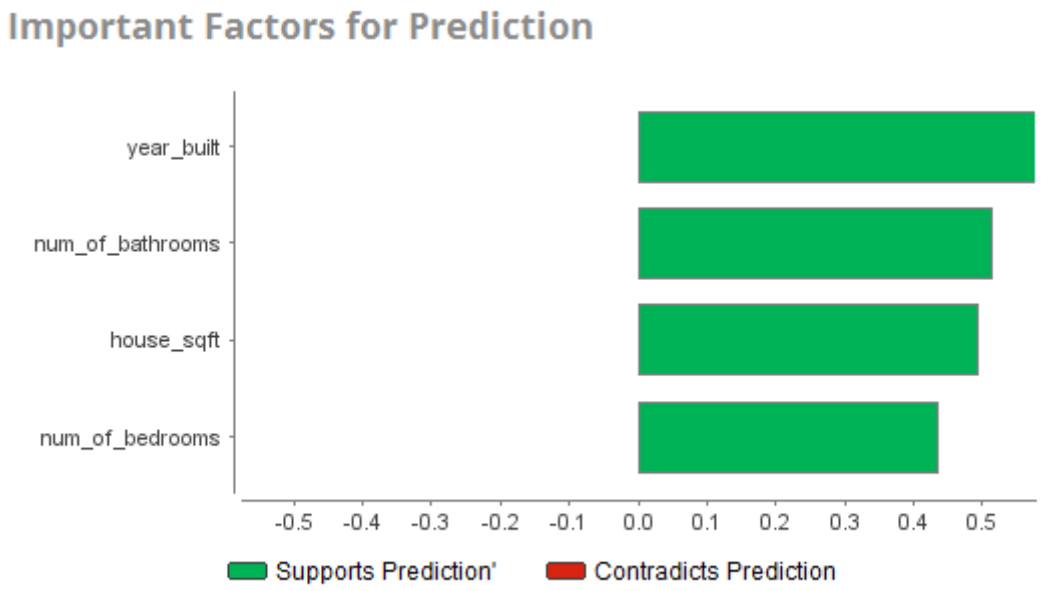
10. Casa con pocos cuartos, construida actualmente y con gran superficie



ANALÍTICA DE DATOS

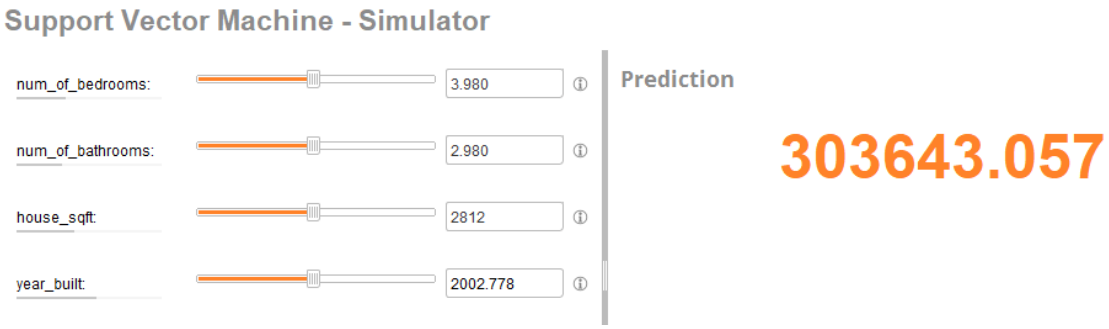
F. Algoritmo: Support Vector Machine

Factores:

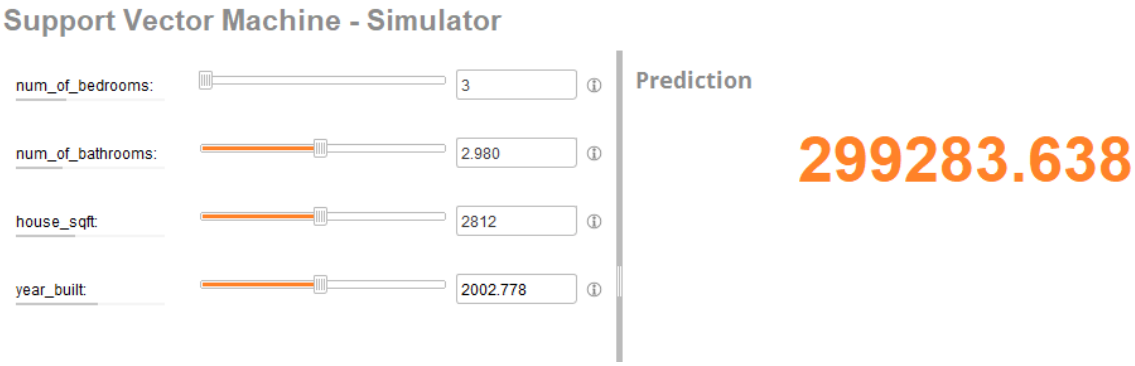


Escenarios:

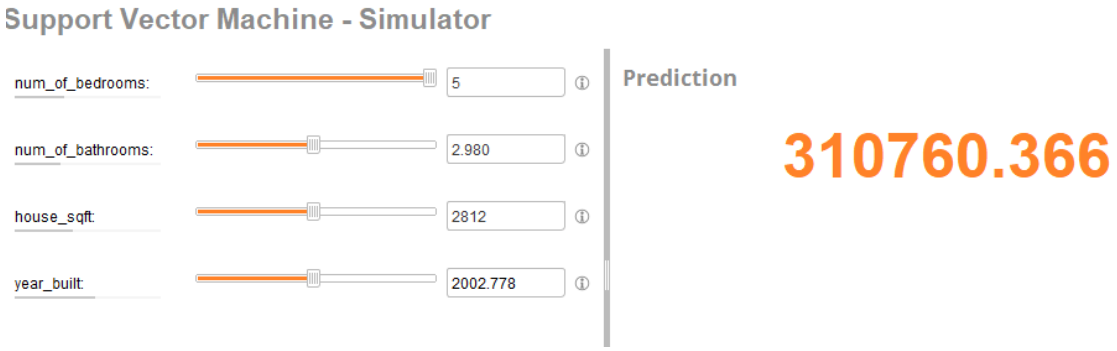
1. Con número promedio de datos



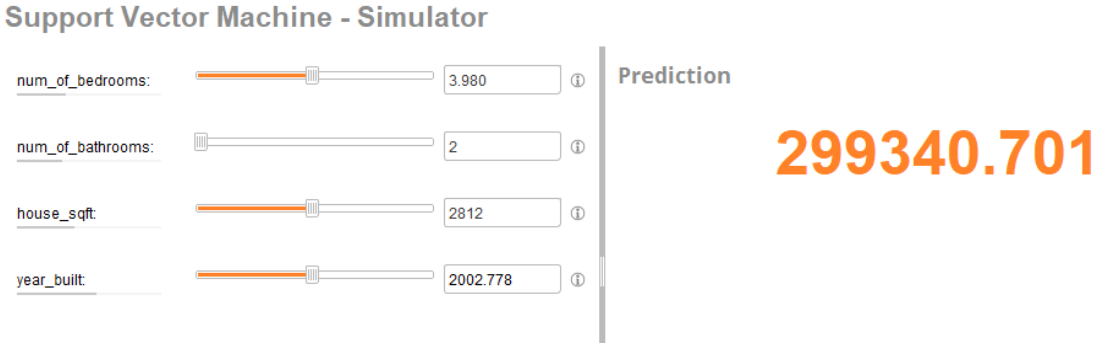
2. Con bajo número de cuartos



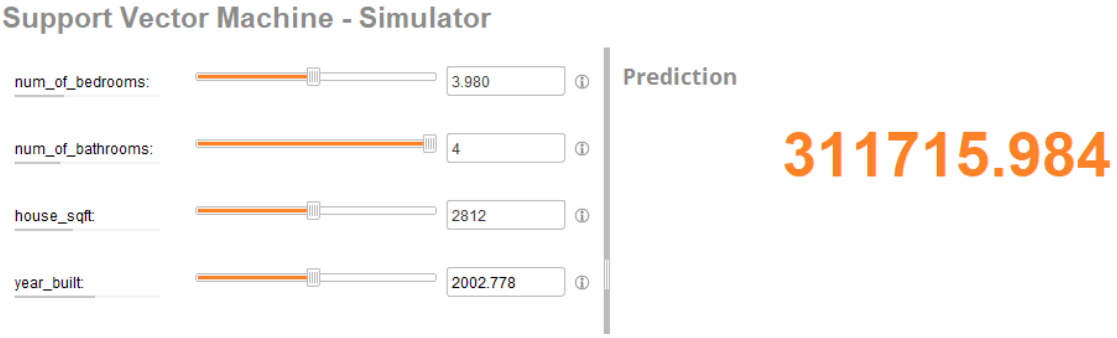
3. Con alto número de cuartos



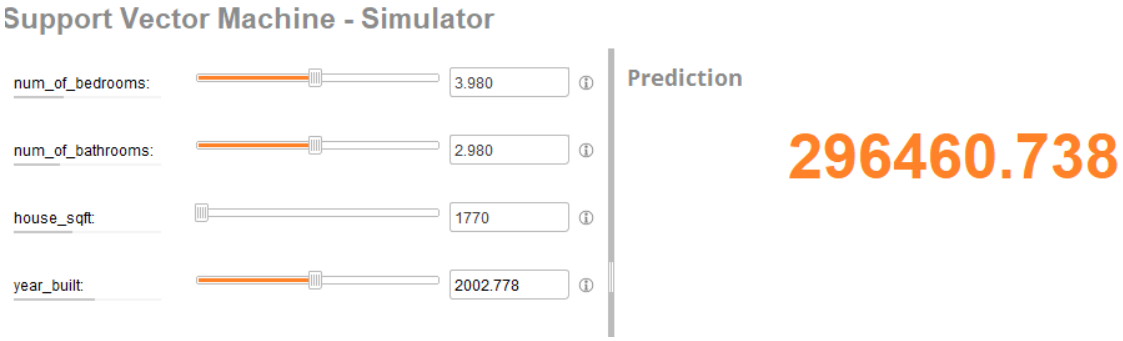
4. Con bajo número de baños



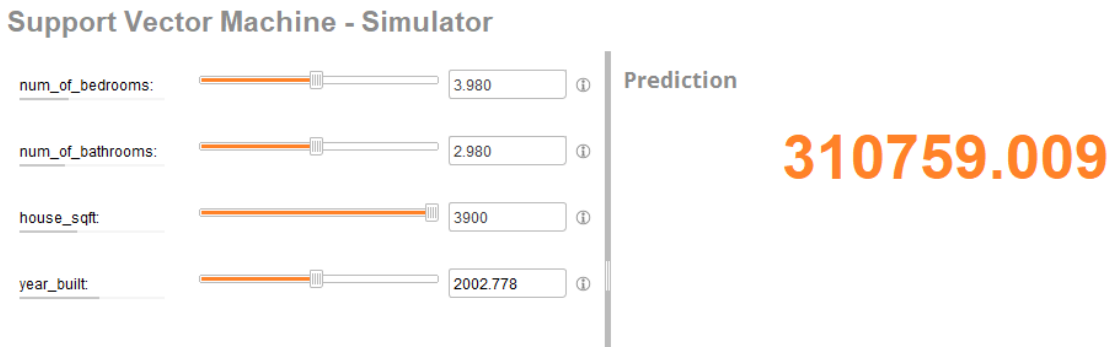
5. Con alto número de baños



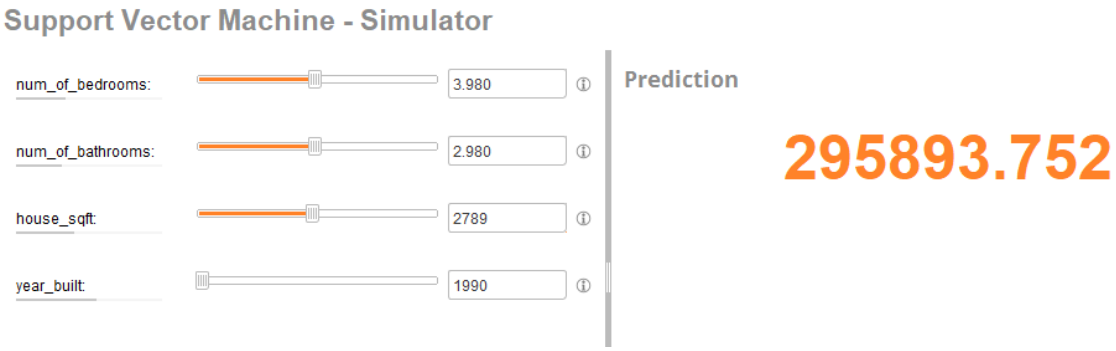
6. Con poca superficie



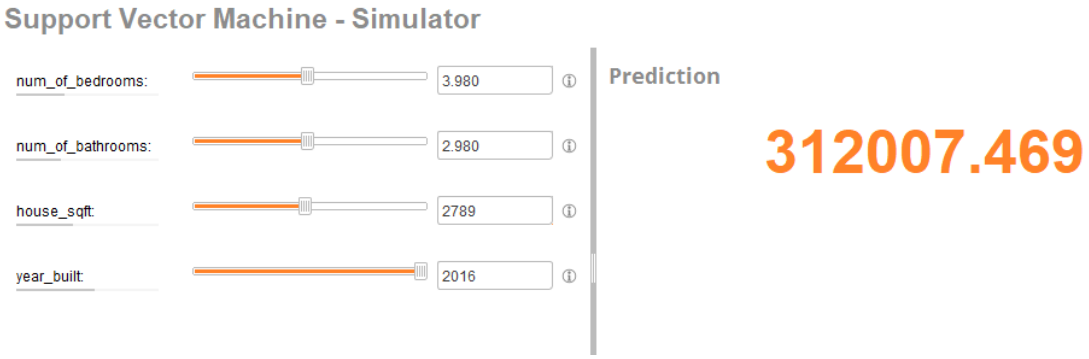
7. Con gran superficie



8. Casa antigua



9. Casa moderna



10. Casa con pocos cuartos, construida actualmente y con gran superficie