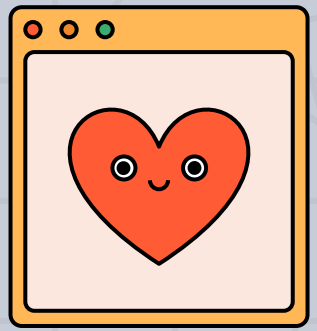
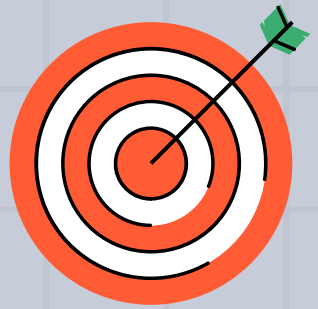
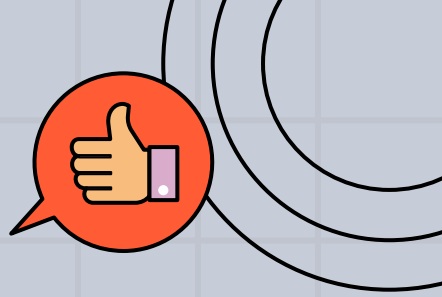
Prediction Model for Sillicon Valley Real Estate

By Indira



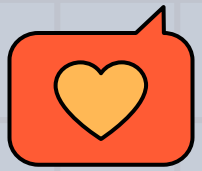


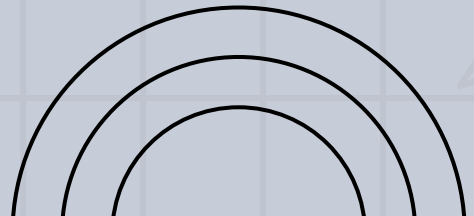
Context



I'm an AI development consultant at a Silicon Valley startup that provides real estate investment services. Relationship managers mentioned that demand has increased recently and it is becoming difficult to do personalized estimates. As a result, the company commissioned me to automate this task with a predictive model.

For this, I retrieved a database that contains median house prices for California neighborhoods from the 1990 census





Methodology

01 Dataset discovery
16512 rows and 11 columns

02 The basics
The “unnamed” column is removed

03 Null values
176 in “total_bedrooms” column

04 The duplicates
None discovered

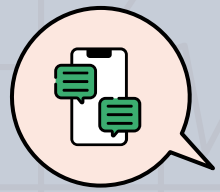
05 Categorical values
Encoding “ocean_proximity”

06 Sum up
Some stats and Data Viz

My models

Linear Regression

Random Forest

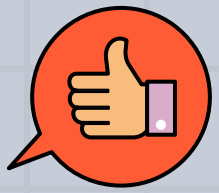


Dummy Regressor

Those seemed to be the best since they can predict a continuous value such as the median price for real estate

35h 55m 23s

Time “officially” allocated for the project



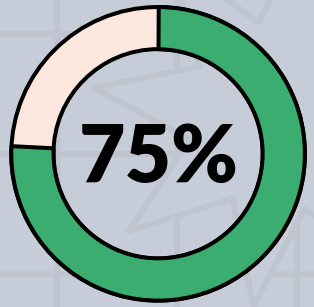
105h00m

Time that should be allocated in the future

386,000

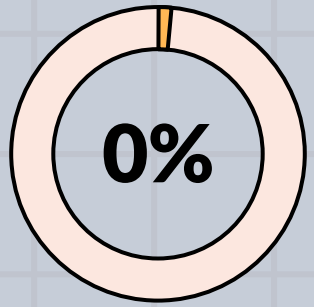
Times, I wished to transfuse coffee in my blood

Key elements



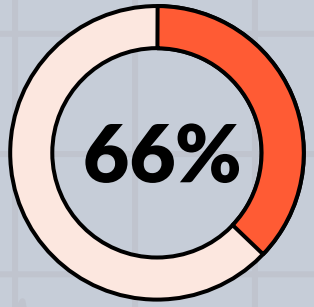
Random Forrest

The best model for me in this situation but sadly not the preferred one



Dummy regressor

A good way to have a baseline model



Linear regression

Actually had a good result but not over the 70% I hoped for...

Thanks

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**