

# **Predictive Analytics: Regression**

</talentlabs>

## Overview

This case study aims to give an idea of applying Machine Learning to predict prices. As you know, regression is a technique for investigating the relationship between independent features and an outcome. In ML, regression is used for predictive modelling.

Here, you will be developing a basic machine learning regression model to predict house price based on the features provided.

## Problem Statement

As a junior data analyst in a real estate company, It is your job to predict the sales price for each house. For each Id in the test set, you must predict the value of the SalePrice variable. Your project manager asked you to proceed with the dataset provided. In this case study, you need to develop an ML model that predicts the house price based on the features provided in the dataset.

## Dataset

You can download the dataset and understand the metadata (the data definition) from the official Kaggle page at <https://www.kaggle.com/competitions/home-data-for-ml-course/data>.

Please download the data from this path for this project. Note that there is a train.csv and test.csv file. You will need to build a model with train.csv and use test.csv to test your model.

## Expected Outcomes

Create a machine learning regression model that predicts the "price" of the property with the given features.





- Explore the data set and do data cleaning if needed.

- Do exploratory data analysis to understand the data.
- Set your target variable here, “price” and understand how the variables are related to the target variable.
- Since the features are low here, use feature engineering to add the features.
- Work on various ML algorithms for regression to select an ML model that performs best for the data at hand.
- Evaluate your results using performance metrics and find the best algorithm that predicts the house price. You should use “Root-Mean-Squared-Error (RMSE) between the logarithm of the predicted value and the logarithm of the observed sales price” as the performance metrics.
- To evaluate your model, you can submit your predictions to the Kaggle platform here: <https://www.kaggle.com/competitions/home-data-for-ml-course/overview/evaluation>. All you need to do is to export your prediction in the submission file format as described on the page, and upload by clicking the “Submit Predictions” button.

**Note:** You need to submit one/two Ipython notebooks which clearly explains the thought process behind your ML model (either in comments or markdown text), code and relevant plots. Also, all the visualizations and plots must be done in Python (should be present in the Ipython notebook).

Please include a proof that you achieve a RMSE > 15400 on the Kaggle platform.

Sample screenshot as below:

2884	Talha Anwer		15402.86549	1	2y
2885			15403.61256	1	9m
 Your First Entry! Welcome to the leaderboard!					
2886	Ashoka98228		15403.88393	8	1y

**Thank you.**

**All The Best :)**