

# Data Intake Report

Name: Deployment on Cloud: California House Pricing Dataset

Report date: 03/06/2023

Internship Batch: LISUM21

Version: 1.0

Data intake by: Zohra Bouchamaoui

Data intake reviewer:

Data storage location: GitHub

## Tabular data details:

<b>Total number of observations</b>	20640
<b>Total number of files</b>	0
<b>Total number of features</b>	8
<b>Base format of the file</b>	Scikit-learn package
<b>Size of the data</b>	N/A

## Proposed Approach:

- No missing attribute values.
- We assume a linear relationship between the features and the target variable, since our model is a regression.
- We assume independence among the data points in the dataset. This means that this assumption implied that the observations are not influenced by each other and there is not autocorrelation present in the data.
- We assume that the predictor variables are not highly correlated with each other since multicollinearity can lead to unstable and unreliable estimates of the model's coefficients.
- We assume there are no extreme outliers.