

SDAIA ACADEMY

T5 Boot Camp

LINEAR REGRESSION PRESENTATION

Supervised by: Mr. Mohamed El-Deeb

Linear Regression module

AGENDA

- Goal
- Data Collection
- EDA
- Data Cleaning
- Pre-Processing
 - Feature preparation
 - Feature Engineering
 - Feature Selection
- Modeling
- Models Evaluation

GOAL

• Predicting the Discount Percentage.

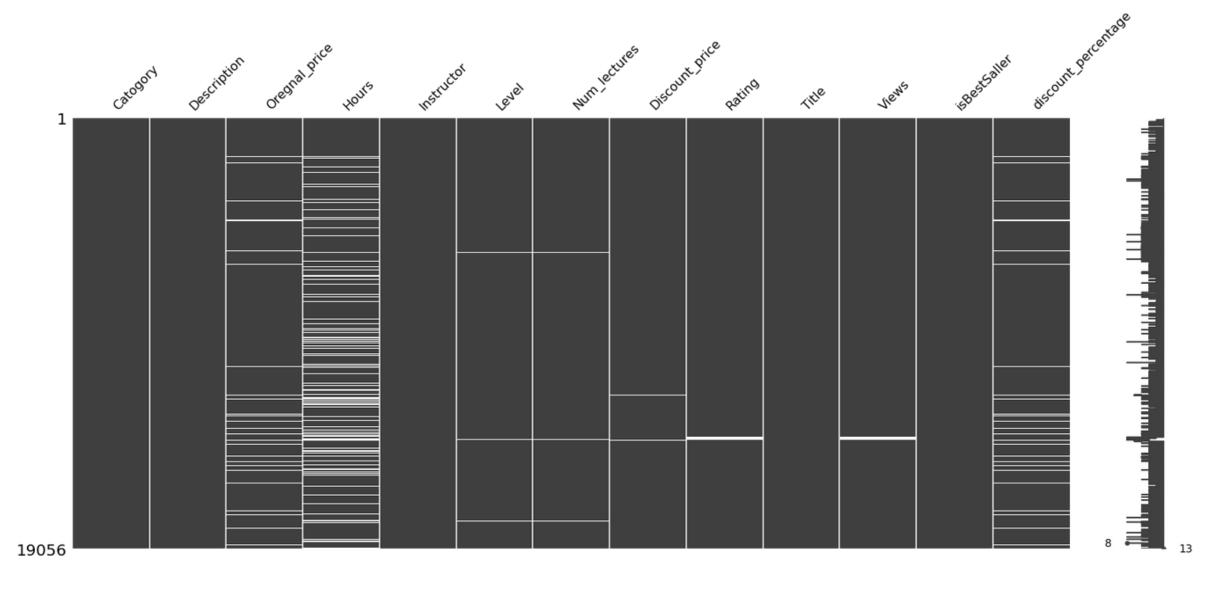


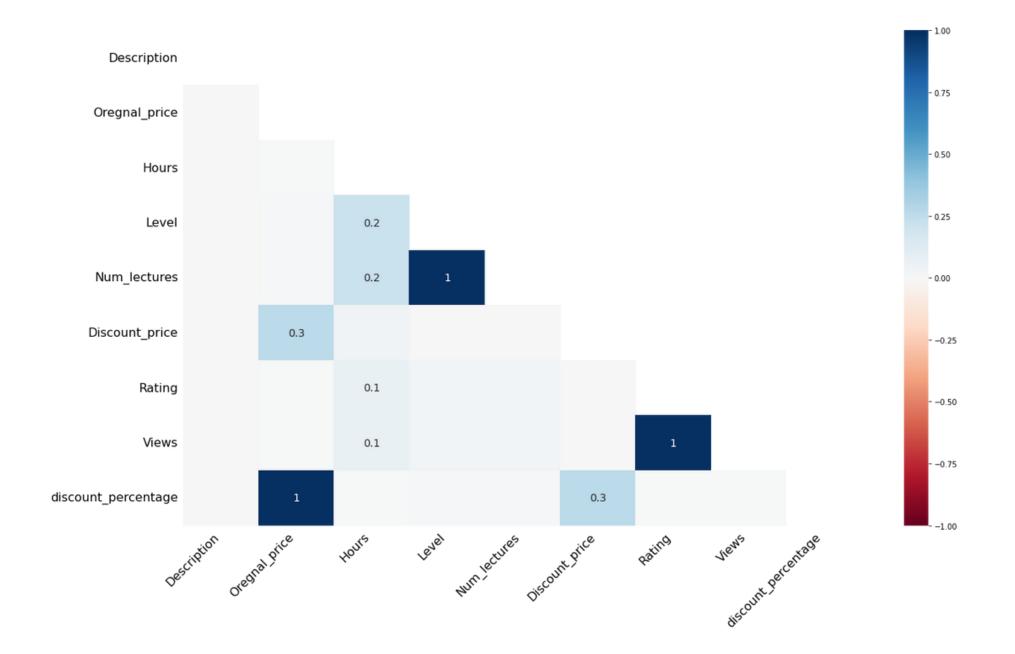
DATA COLLECTION

- Web Scribing "Udemy"
- 12 Features.
- Almost 20K Record.

EDA

• Null Values

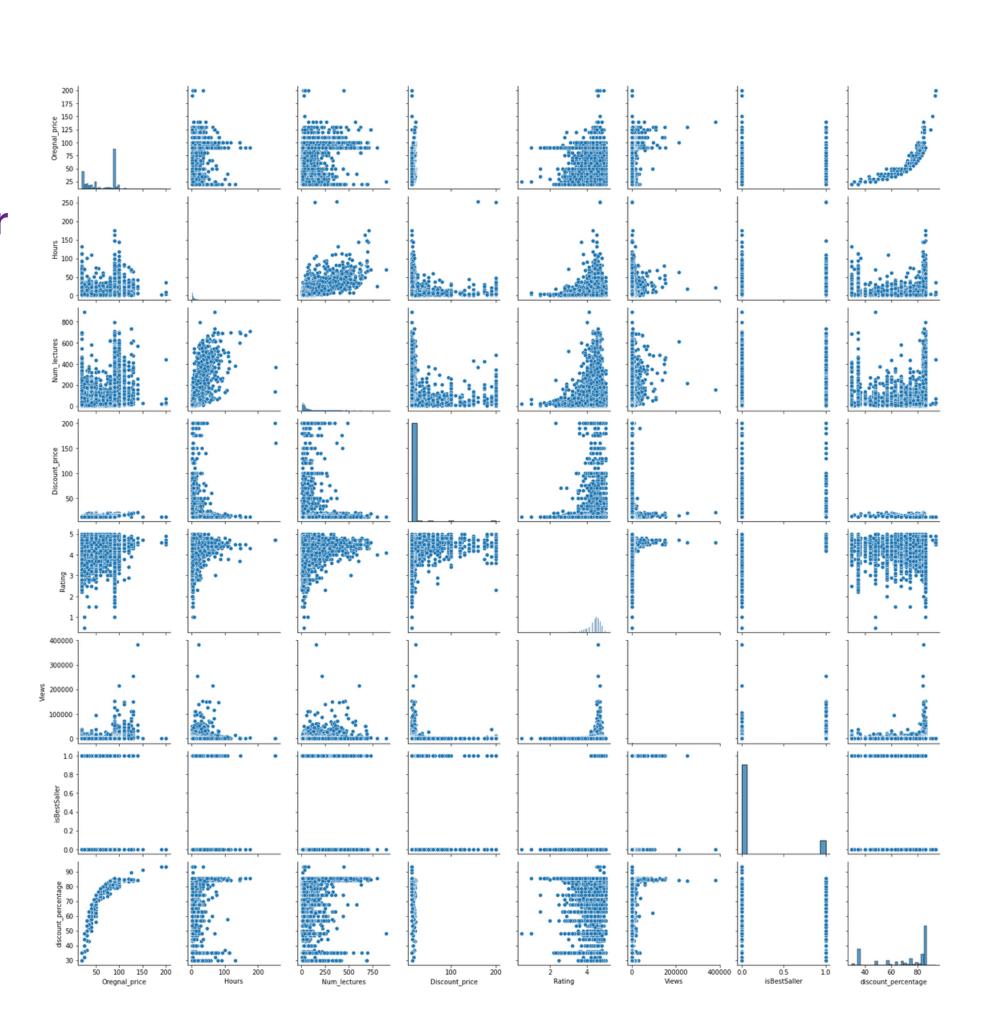






EDA

• Corr



DATA CLEANING

Null Handling

> Drop 4K

Outlier Handling

no Outlier

PRE-PROCESSING

Feature preparation

Categorical into Dummy

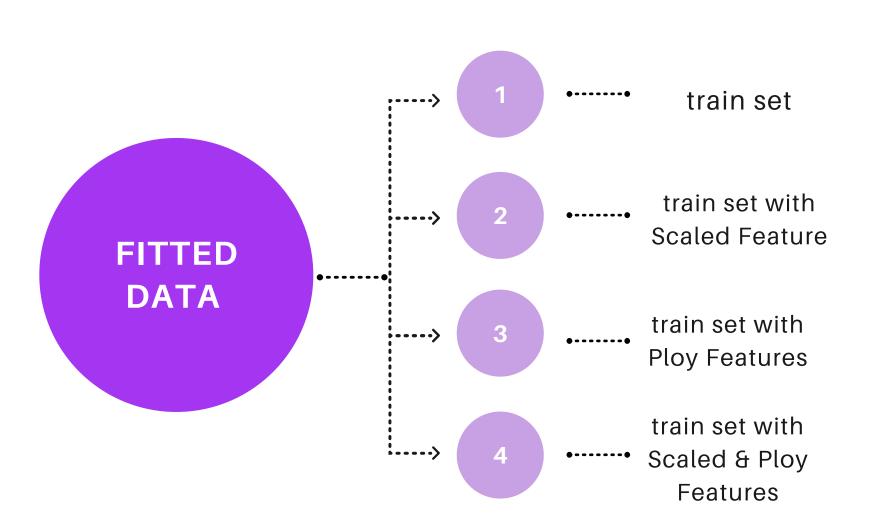
Feature Engineering

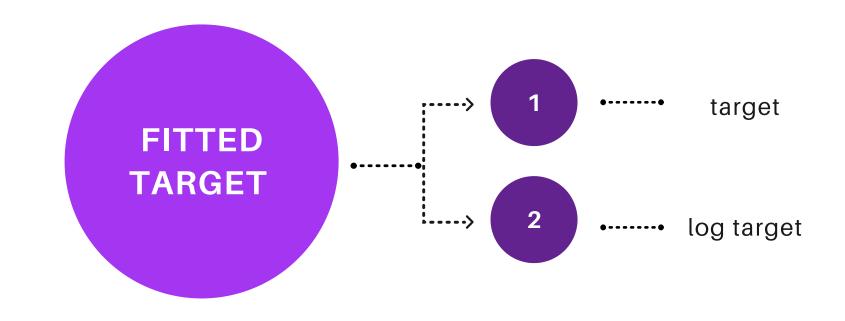
Scale Ploy

Feature Selection

LASSO

MODELING





MODELS EVALUATION Score train validation 0.88307 0.87831 target train set 0.81354 0.8057 • log target 0.88307 0.87831 target train set with Scaled Feature 0.81354 0.8057 •---- log target **FITTED DATA** 0.88384 0.8784 target train set with 3 Ploy Features 0.80565 0.81444 •---- log target train set with 0.88384 0.8784 target Scaled & Ploy •----Features 0.80565 0.81444 •---- log target

MODELS EVALUATION Score train validation 0.88307 0.87831 target train set 0.81354 0.8057 • log target 0.88307 0.87831 target train set with Scaled Feature 0.81354 • log target 0.8057 **FITTED DATA** 0.88384 0.8784 target train set with 3 Ploy Features 0.80565 0.81444 • log target train set with 0.8784 0.88384 target Scaled & Ploy •----Features 0.80565 0.81444 •---- log target

MODELS EVALUATION

Baseline vs.
Best experiment



only numerical Features target

	Score		
	train		validation
0.88116			0.87821



train set with
Ploy Features
target

	Score		
train	validation		
0.88384	0.8784		

