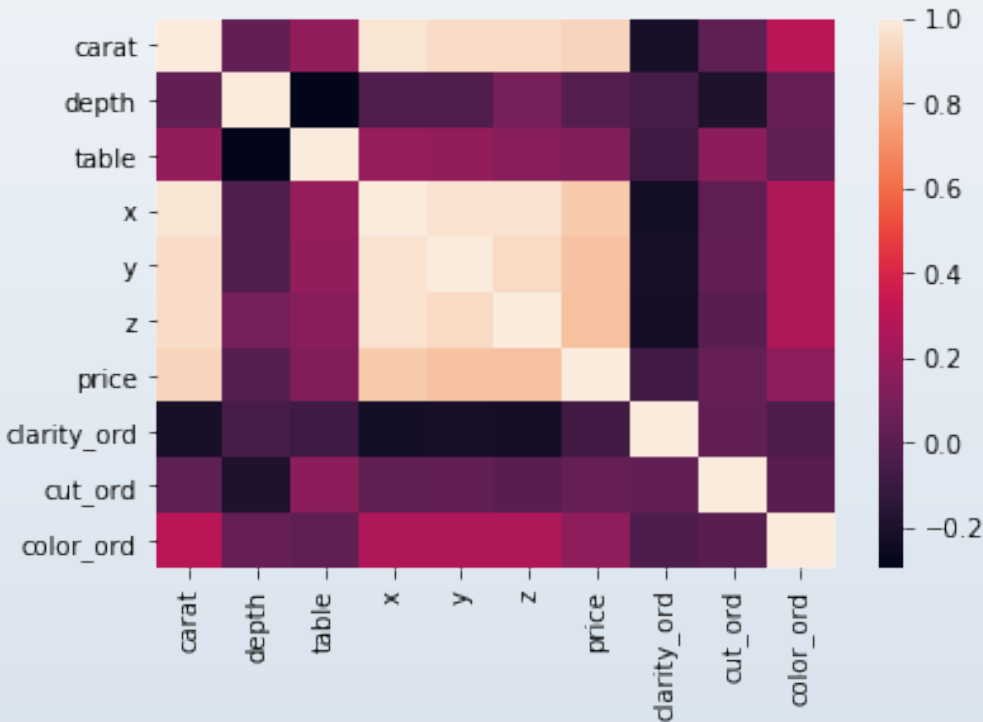
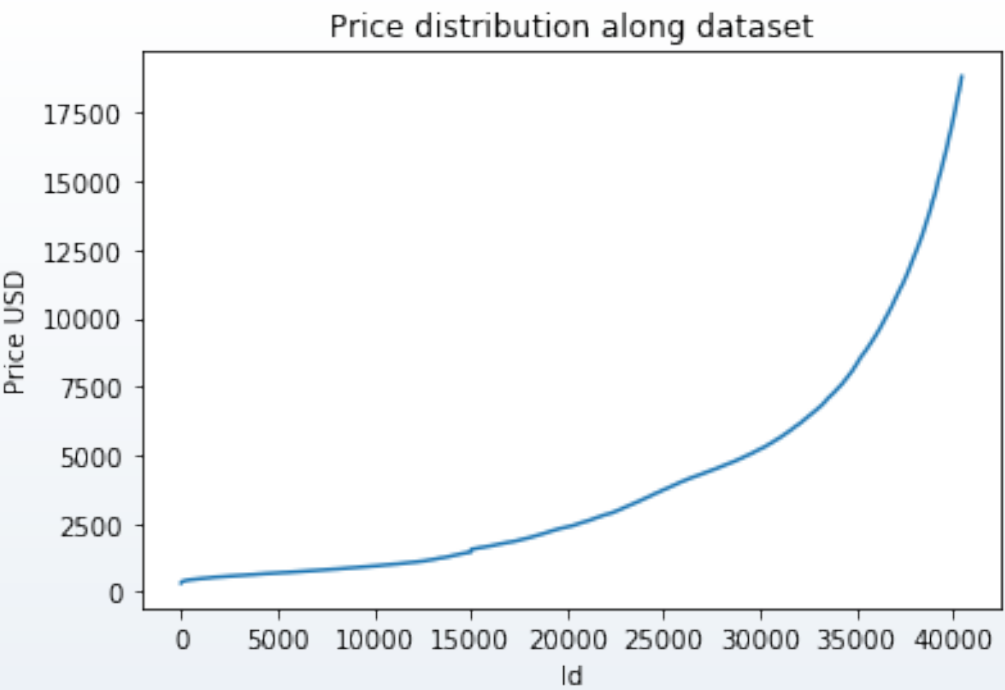


## Data Analysis

- Categorical → ordinal\_values (get\_dummies)
- x,y,z Possilbe overfitting?
- Distinct models for less/more expensive diamonds?



## Best iterations

Including x,y,z  
Models trained for 100% data

**Sklearn:**  
Kfolds: 10 (cross-validation)  
Hyperparameters:  
n\_iterators: 400  
max\_depth:35

**H2O:**  
AutoML (200s)

## Results

| Position | Model                    | RMSE      |           | R2     |        |
|----------|--------------------------|-----------|-----------|--------|--------|
|          |                          | Train     | Test      | Train  | Test   |
| 1        | H2OStacked Ensemble      | 365.1289  | 535.7720  | 0.9917 | 0.9820 |
| 2        | LGBMRegressor            | 551.3147  | 626.3562  | 0.9808 | 0.9758 |
| 3        | LGBMRegressor            | 546.3852  | 640.2655  | 0.9814 | 0.9729 |
| 4        | RandomForestRegressor    | 567.9414  | 636.3598  | 0.9797 | 0.9741 |
| n        | H2ODeepLearningEstimator | 646.5333  | 660.0752  | -      | -      |
| n        | LinearRegression         | 1386.2388 | 1376.4495 | 0.8799 | 0.8793 |
| n        | ElasticNet               | 2973.7926 | 2970.1369 | 0.4487 | 0.4446 |
| n        | KNeighborsRegressor      | 1999.5835 | 2478.3588 | 0.7488 | 0.6159 |
| n        | DecisionTreeRegressor    | 757.9989  | 872.8740  | 0.9640 | 0.9515 |

