



Missing Data Imputation with Feature-Engine

Feature-Engine



- <https://www.trainindata.com/feature-engine>
- <https://feature-engine.readthedocs.io/en/latest/>
- https://github.com/solegalli/feature_engine

pip install feature-engine

Feature-Engine: Advantages

- Feature-Engine includes all the feature engineering techniques described in the course
- Feature-Engine works like to Scikit-learn, so it is easy to learn
- Feature-Engine allows you to implement specific engineering steps to specific feature subsets
- Feature-Engine can be integrated with the Scikit-learn pipeline allowing for smooth model building
- **Feature-Engine allows you to design and store a feature engineering pipeline with bespoke procedures for different variable groups.**

Feature-Engine: with fit and transform



- `fit()` → learns parameters from train set
- `transform()` → transforms data

Feature-Engine transforms specific variables



Feature engine allows you to specify variable groups easily

```
1  # Let's do mean imputation this time
2  # and let's do it over 2 of the 3 numerical variables
3
4  imputer = mdi.MeanMedianImputer(imputation_method='mean',
5                                  variables=['LotFrontage', 'MasVnrArea'])
6
7  imputer.fit(X_train)
```

```
MeanMedianImputer(imputation_method='mean',
                   variables=['LotFrontage', 'MasVnrArea'])
```

Feature-Engine stores the transformation parameters



```
| 1 # now the imputer uses only the variables we indicated  
| 2  
| 3 imputer.variables
```

```
['LotFrontage', 'MasVnrArea']
```

```
| 1 # and we can see the value assigned to each variable  
| 2 imputer.imputer_dict_
```

```
{'LotFrontage': 69.66866746698679, 'MasVnrArea': 103.55358898721731}
```



Feature-Engine: Limitations

✓ I would like to hear your feedback!!!

THANK YOU

www.trainindata.com