Train In Data

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

```
# now the imputer uses only the variables we indicated

imputer.variables
```

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

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

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

Feature-engine: Limitations

- ✓ I would like to hear your feedback!!!

Train In Data

THANK YOU

www.trainindata.com