




Missing Data Imputation with Feature-engine

Feature-engine



Feature-engine is an open-source Python library with multiple transformers to engineer and select features to use in machine learning models.



What is unique about Feature-engine?



Feature-engine exhaustive battery of feature transformations.



Feature-engine and feature subsets

- Transformations can be applied to a feature subset.
- No need to slice data or use additional classes.

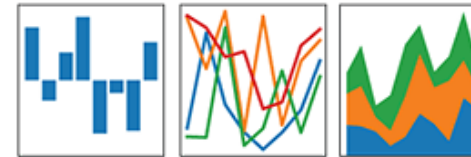


Feature-engine works with dataframes

- Takes in and returns a Pandas dataframe
 - data analysis + deployment

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

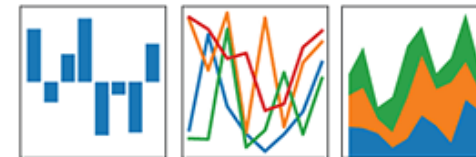


Feature-engine and variables

Automatically recognises numerical, categorical and datetime variables.

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Feature-engine: fit() and transform()



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

Feature-engine



- <https://feature-engine.readthedocs.io/en/latest/index.html>
- https://github.com/feature-engine/feature_engine

`pip install feature-engine`

`conda install -c conda-forge feature-engine`

Feature-engine

- Allows application of each feature engineering technique to a **subset of features**

```
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

```
2  
3 imputer.variables
```

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

```
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