Categorical encoding - definition

Categorical encoding refers to replacing the category
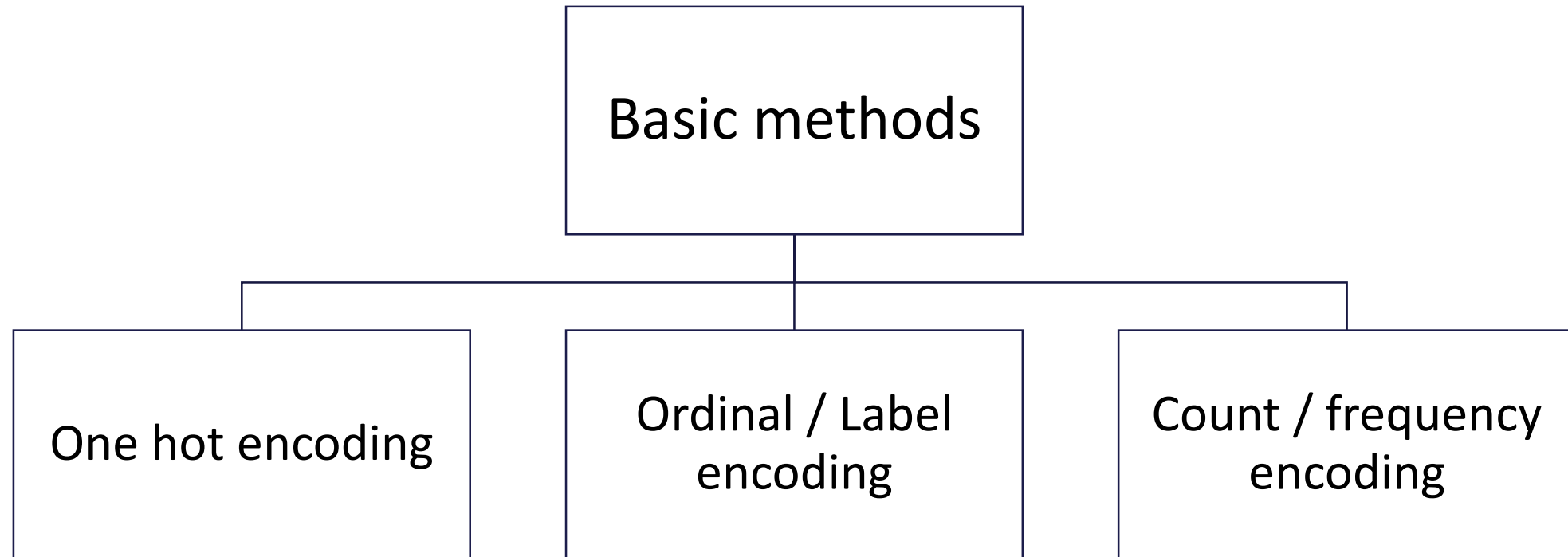
strings by a numerical representation.
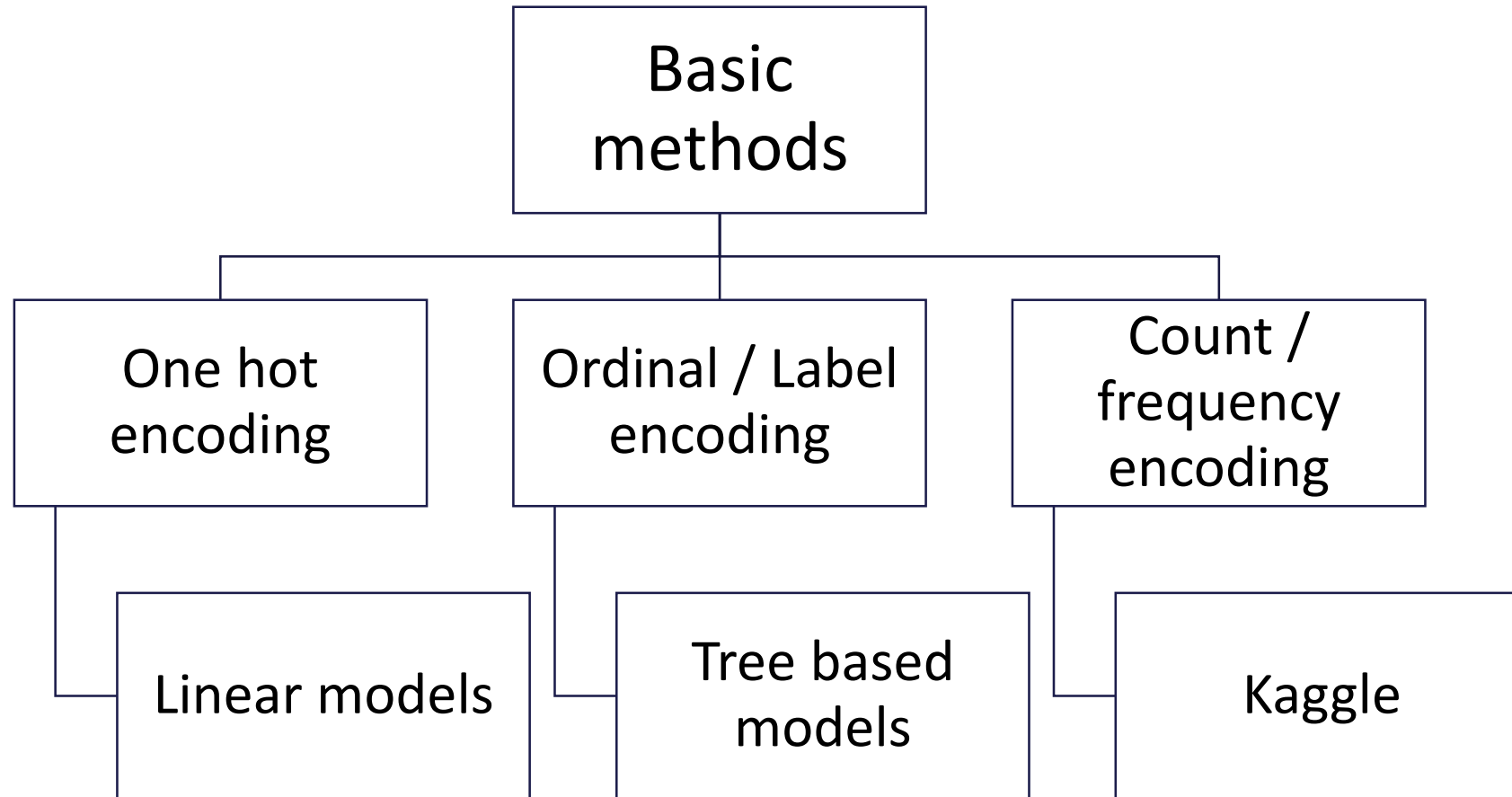
Categorical encoding - goal

- To produce variables that can be used to train machine learning models.

- To build predictive features from categories.
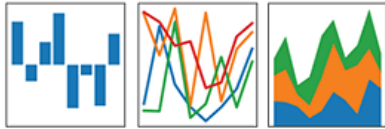
Basic categorical encoding methods

```
                    ┌─────────────────────┐
                    │                     │
                    │   Basic methods     │
                    │                     │
                    └──────────┬──────────┘
            ┌──────────────────┼──────────────────┐
┌───────────┴─────────┐ ┌──────┴──────────┐ ┌─────┴────────────┐
│                     │ │                 │ │                  │
│  One hot encoding   │ │ Ordinal / Label │ │ Count / frequency│
│                     │ │    encoding     │ │    encoding      │
└─────────────────────┘ └─────────────────┘ └──────────────────┘
```

Basic categorical encoding methods

Categorical encoding in Python

pandas

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

scikit learn

Feature-Engine

Category Encoders

Train In Data

Objectives

Learn multiple techniques

Understand their impact on the variable and the machine learning model

Learn how to implement it with pandas, Scikit-learn, Feature-Engine, and Category encoders

Train In Data

Content

- Presentation and video

  - Explanation of the technique

- Accompanying Jupyter notebook

  - Implementation in pandas

  - Implementation in Scikit-learn

  - Implementation in Feature-engine

  - Implementation in Category encoders

THANK YOU

www.trainindata.com

Train In Data