



Data Preparation Exam Tips



Brock Tubre
INSTRUCTOR

Data Preparation

- Know what data preparation is and why it is important in Machine Learning.
- Understand the different techniques used for preparing data.

Categorical Encoding

- Know why categorical encoding is used for certain ML algorithms.
- Understand the difference between ordinal and nominal categorical features.
- Understand that categorical data is qualitative and continuous data is quantitative.
- Know what one-hot encoding is and when to use it.

Numeric Feature Engineering

- Know what numeric feature engineering is and why it is important.
- Know different techniques used for feature engineering numeric data.
- Know the different types of feature scaling and when they should be used.
 - Normalization
 - Standardization
- Know what binning is and when it should be used.

Text Feature Engineering

- Know what text feature engineering is and why it is important.
- Know different techniques used for feature engineering text data.
 - N-Gram
 - Orthogonal Sparse Bigram (OSB)
 - Term Frequency-Inverse Document Frequency (tf-idf)
 - Removing punctuation
 - Lowercase transformation
 - Cartesian product
- Understanding why feature engineering dates is important.
- Know the questions we can answer when dates are transformed.

Other Feature Engineering

- Know other types of feature engineering covered.

Handling Missing Values

- Know why handling missing values is an important step in data preparation.
- Know the different techniques used for handling missing values.
- Understand implications of dropping rows.
- Understand what data imputation is.

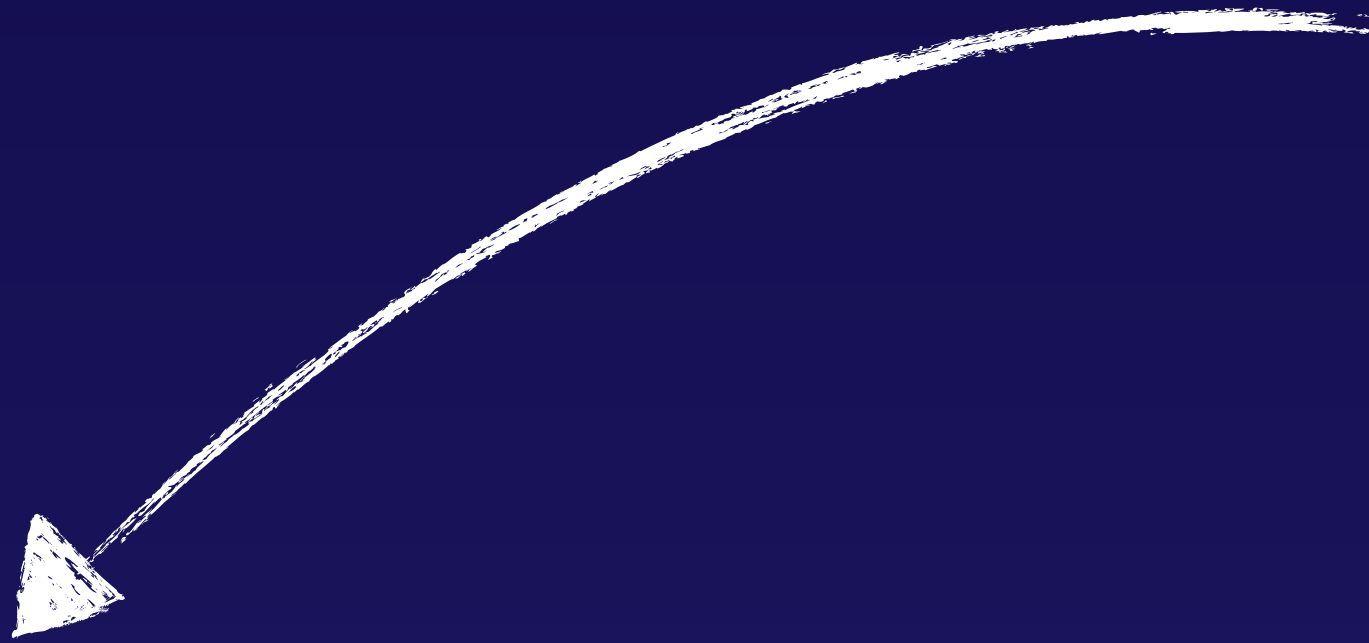
Feature Selection

- Know what feature selection is and why it is important.
- Understand the difference in feature selection and Principle Component Analysis (PCA).

Data Preparation Tools

- Know the different AWS services that allow you to transform data.
- Know what a Data Catalog, Crawlers, and Jobs are in AWS Glue.
- Be able to identify the different AWS services and when to use one transformation tool over another.

**Be sure to read/watch
the additional resources**



**Be sure to read/watch
the additional resources**

