

Data Preprocessing: Case Study

Download the sample data below to solve this Data Science case study on Data Preprocessing.

In Data Science, the ability to preprocess raw data effectively is crucial for accurate and insightful analysis. The table below consists of both numeric and categorical features. The numeric features, denoted as NumericFeature1 and NumericFeature2, contains missing values represented as NaN. The CategoricalFeature comprises distinct categories represented by alphabetic characters.

+-----+-----+-----+			
NumericFeature1 NumericFeature2 CategoricalFeature			
+-----+-----+-----+			
1.0	7	A	
2.0	8	B	
NaN	9	NaN	
4.0	10	A	
5.0	11	B	
6.0	50	C	
+-----+-----+-----+			

Your task is to create a robust data preprocessing pipeline using Python, capable of handling missing values, standardizing numeric features, and detecting and removing outliers, thus enhancing the overall quality and integrity of the data. The pipeline should encompass a range of preprocessing techniques to ensure the resulting data is of high quality and suitable for subsequent analysis.