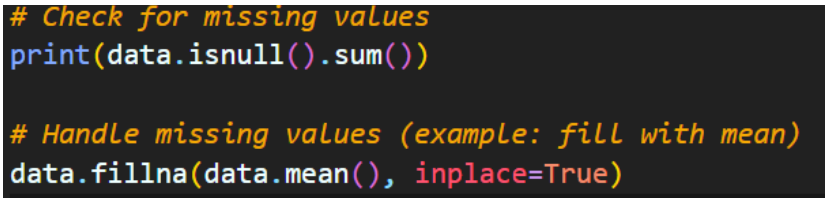
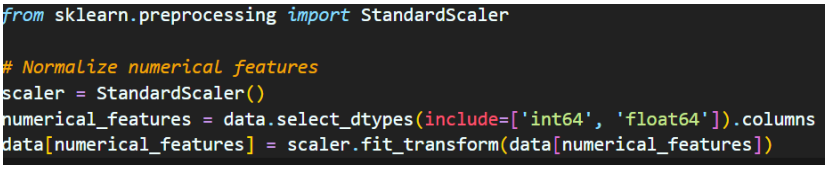
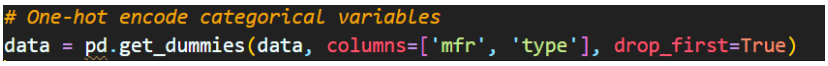
Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720184497
Project Title	Cereal Analysis Based on Ratings by using Machine Learning Techniques
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	Basic statistics, dimensions, and structure of the data.
Univariate Analysis	Exploration of individual variables (mean, median, mode, etc.).
Bivariate Analysis	Relationships between two variables (correlation, scatter plots).
Multivariate Analysis	Patterns and relationships involving multiple variables.
Outliers and Anomalies	Identification and treatment of outliers.
Data Preprocessing Code Screenshots	
Loading Data	`# Load the dataset into the preferred environment` `data = pd.read_csv('cereal.csv')`

Handling Missing Data	
Data Transformation	
Feature Engineering	
Save Processed Data