

```

1 student_transformer = Pipeline(steps=[
2     # Map binary categorical variables to 0/1
3     ('map_gender', CustomMappingTransformer('gender', {'Male': 0, 'Female': 1, 'Other': 2})),
4     ('map_part_time_job', CustomMappingTransformer('part_time_job', {'No': 0, 'Yes': 1})),
5     ('map_extracurricular', CustomMappingTransformer('extracurricular_participation', {'No': 0, 'Yes': 1})),
6
7     # Map ordinal categorical variables (preserve order)
8     ('map_diet_quality', CustomMappingTransformer('diet_quality', {'Poor': 0, 'Fair': 1, 'Good': 2})),
9     ('map_internet_quality', CustomMappingTransformer('internet_quality', {'Poor': 0, 'Average': 1, 'Good': 2})),
10    ('map_parental_education', CustomMappingTransformer('parental_education_level',
11    {'None': 0, 'High School': 1, 'Bachelor': 2, 'Master': 3})),
12
13    # Apply Tukey outlier detection to numeric features
14    ('tukey_age', CustomTukeyTransformer(target_column='age', fence='outer')),
15    ('tukey_study_hours', CustomTukeyTransformer(target_column='study_hours_per_day', fence='outer')),
16    ('tukey_social_media', CustomTukeyTransformer(target_column='social_media_hours', fence='outer')),
17    ('tukey_netflix', CustomTukeyTransformer(target_column='netflix_hours', fence='outer')),
18    ('tukey_attendance', CustomTukeyTransformer(target_column='attendance_percentage', fence='outer')),
19    ('tukey_sleep', CustomTukeyTransformer(target_column='sleep_hours', fence='outer')),
20
21    # Apply robust scaling to numeric features
22    ('scale_age', CustomRobustTransformer(target_column='age')),
23    ('scale_study_hours', CustomRobustTransformer(target_column='study_hours_per_day')),
24    ('scale_social_media', CustomRobustTransformer(target_column='social_media_hours')),
25    ('scale_netflix', CustomRobustTransformer(target_column='netflix_hours')),
26    ('scale_attendance', CustomRobustTransformer(target_column='attendance_percentage')),
27    ('scale_sleep', CustomRobustTransformer(target_column='sleep_hours')),
28
29    # Drop the original exam_score and student_id columns
30    ('drop_columns', CustomDropColumnsTransformer(['exam_score', 'student_id'], action='drop')),
31
32    # Impute any remaining missing values
33    ('impute', CustomKNNTransformer(n_neighbors=5)),
34 ], verbose=True)

```