**NCAA Feature Engineering & Selection**

The final selected features appear robust and highly suitable for effectively answering my research question:

**"Can we predict a player's NCAA division (Division I, II, or III) based on demographic and physical attributes (e.g., height, year/class, state or country of origin, position)?"**

**Strengths of Your Selected Features:**

**1. Division Indicators (Target Variable)**:

* My target (division\_III) clearly reflects your classification objective.  
  *(Note: Ensure your final modeling aligns with predicting the correct division category—Division I, II, or III.)*

**2. Demographic Indicators (year/class)**:

* Features like year\_clean\_freshman, year\_clean\_junior, etc., effectively capture demographic/class attributes, which are important predictors of division differences.

**3. Geographic Attributes**:

* Numerous state features (state\_grouped\_ca, state\_grouped\_fl, state\_grouped\_international, etc.) capture geographic recruitment patterns.
* Presence of the state\_grouped\_international feature is beneficial to explore international recruitment effects on NCAA division membership.

**4. Player Positions**:

* Position features (position\_grouped\_forward, position\_grouped\_guard) directly reflect team composition strategies related to NCAA divisions.

**5. Physical Attributes (Height)**:

* taller\_than\_avg (binary) and total\_inches\_scaled (standardized continuous feature) provide strong numeric indicators for player physical characteristics crucial in athletic prediction.