Strategy for Matching No-Bid Sales to Actual Harvests

PROMPT: help me develop a strategy to match the no bid observations to subsequent sale that led to an actual harvest. actual harvests are recoded in the timber harvest dataset. when a tract of timber land needs to be harvested, forest service managers create a sale controct that loggers can bid on. when a contract recieves no bids, the fs manager revises the contract to improve its chances of being sold. nearly all timber tracts are eventually sold. this means that each observation in the no bids dataset is related to an actual sale from the timber harvest dataset. the challenge is that there is no unique key that directly links the two datasets.

I'll outline a comprehensive approach to match observations in the no_bids_bienville.csv dataset with successful sales in the TimberHarvestBienville shapefile.

Challenge Understanding

When a timber sale receives no bids, it's typically revised and reoffered until it sells. Our goal is to link each no-bid record to its eventual successful sale despite lacking a direct key between datasets.

Matching Strategy

1. Administrative Filter

- Match on administrative codes (FORESTORGC, DISTRICTOR, forest_code, district_code)

2. Temporal Filter

- Dates should match directly or at least be very close. Try a window of time.
- TIM dates could be incorrect due to data entry issues.
- I would use this as a "soft" filter

3. Area Matching

- Compare gross_sale_area in no_bids_bienville with GIS_ACRES in TimberHarvestBienville
- Compare volume metrics (volume_mbf, volume_ccf, sale_volume_mbf, sale_volume_ccf) with volumes in the harvest data
- Similar approach to the Temporal Filtering because there may be measurement errors

4. Scoring System Implementation

Implement a weighted scoring system:

1. Assign weights to each matching criterion (spatial match might get highest weight)
2. Calculate a composite score for each potential no-bid/harvest pair
3. Set a threshold score for determining matches