✅ **Summary of Findings: Parkinson's Disease Dataset EDA**

1. **Target Distribution**:
    - The dataset is slightly imbalanced with more patients diagnosed with Parkinson's (status=1) than healthy individuals (status=0).
    - Class distribution suggests a need to account for imbalance if used in modeling.
2. **Data Structure & Integrity**:
    - The dataset contains 195 samples and 24 features.
    - No missing values were detected.
    - Most features are continuous and derived from voice measurements.
3. **Correlation Analysis**:
    - Strong inter-feature correlations suggest multicollinearity (e.g., MDVP:Fo(Hz), MDVP:Flo(Hz)).
    - The target variable (status) shows highest correlation with spread1, PPE, spread2, DFA, and Shimmer:APQ5.
4. **Feature Distributions**:
    - Histograms reveal that several features are right-skewed (e.g., PPE, Shimmer).
    - Healthy vs. Parkinson's feature distributions are clearly distinguishable, especially for key features.
5. **Boxplots & Violin Plots**:
    - Clear separation of feature values for the two classes (status=0 and 1) in top features.
    - PPE, spread1, and spread2 show particularly strong separation and low overlap, making them good predictors.
6. **Pairplot Insights**:
    - Visual clustering between healthy and affected groups is apparent in selected features.
    - Linear separability observed in some combinations like spread1 vs PPE.
7. **Statistical Summary**:

- Many features show moderate to high skewness and kurtosis, indicating non-normal distributions.
- Features such as PPE and DFA have heavier tails.

8. **T-Test Results**:
    - Statistically significant differences ($p < 0.05$) were observed between healthy and affected groups for all top correlated features.
    - This statistically validates that the selected features are meaningful indicators of disease status.

🧠 **Key Predictive Features:**

- **Highly Correlated with Parkinson's (status = 1):**
    - spread1, PPE, spread2, DFA, Shimmer:APQ5, Shimmer:DDA
- These features should be prioritized in any machine learning or predictive modeling task.

🔍 **Conclusion:**

- The dataset offers rich, well-separated features that effectively distinguish between Parkinson's patients and healthy individuals.
- Visual and statistical exploration supports the reliability of features like spread1, PPE, and DFA as robust indicators.
- Further modeling efforts (e.g., classification) can be confidently built upon this EDA foundation.