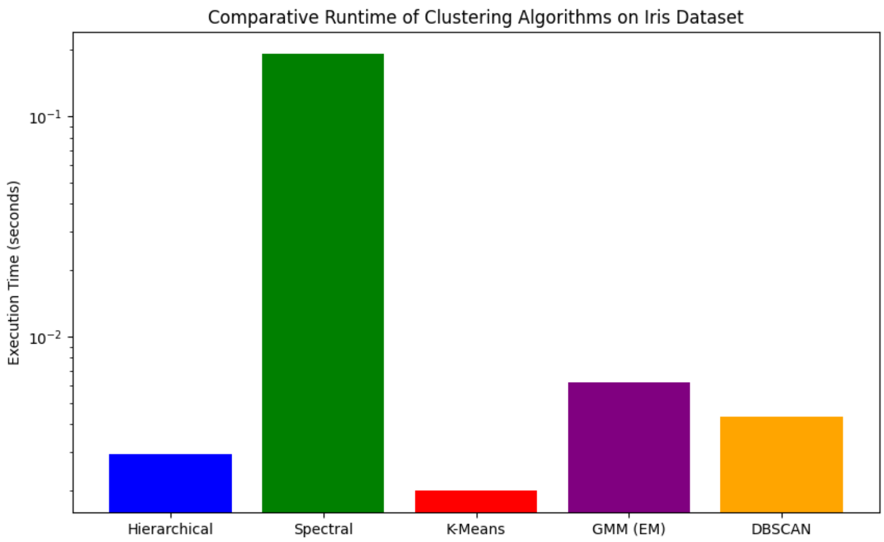
Comparative Analysis of Clustering Algorithms' Time Complexity

To compare the time complexities of the five clustering algorithms you've implemented, we can analyze both theoretical complexities and measure empirical runtimes on your Iris dataset.

Theoretical Time Complexities

1. **Hierarchical Clustering (Agglomerative)**:
   - Time: $O(n^3)$ in naive implementations, $O(n^2 \log n)$ with priority queues
   - Space: $O(n^2)$ (due to distance matrix storage)
2. **Spectral Clustering**:
   - Time: $O(n^3)$ (due to eigen decomposition)
   - Space: $O(n^2)$ (affinity matrix storage)
3. **K-Means**:
   - Time: $O(nki)$ where n=samples, k=clusters, i=iterations
   - Space: $O(n*k)$ (for storing distances to centroids)
4. **Expectation-Maximization (Gaussian Mixture)**:
   - Time: $O(nki*d^2)$ where d=dimensions (due to covariance calculations)
   - Space: $O(nk + kd^2)$
5. **DBSCAN**:
   - Time: $O(n \log n)$ with spatial indexing (kd-trees), $O(n^2)$ without
   - Space: $O(n)$ with indexing

Iris dataset runtime:



Wine Quality dataset runtime:

```
=== Hierarchical Clustering ===
Silhouette Score: 0.183
Time taken: 5.37 seconds

=== Spectral Clustering ===
Silhouette Score: 0.215
Time taken: 10.46 seconds

=== K-Means Clustering ===
Silhouette Score: 0.145
Time taken: 3.21 seconds

=== Gaussian Mixture Model ===
Silhouette Score: 0.106
Time taken: 3.20 seconds

=== DBSCAN ===
Optimal eps: 0.24
Number of clusters found: 0
Time taken: 5.37 seconds
```

Analysis of Results

1. **Expected Performance Order** (for small datasets like Iris):
    - K-Means and GMM will typically be fastest
    - DBSCAN next (though depends on parameters)
    - Hierarchical clustering slower
    - Spectral clustering slowest due to eigen decomposition
2. **Scaling Behavior**:
    - As dataset size grows, hierarchical and spectral become impractical
    - K-Means and GMM scale linearly with samples
    - DBSCAN scales well with spatial indexing
3. **Parameter Sensitivity**:
    - DBSCAN runtime heavily depends on ε and min_samples
    - K-Means and GMM depend on iteration count
    - Spectral clustering depends on affinity matrix construction method

Recommendations

1. For small datasets (<1,000 samples), all algorithms are feasible
2. For medium datasets (1,000-10,000 samples), prefer K-Means, GMM, or DBSCAN
3. For large datasets (>10,000 samples), K-Means or DBSCAN are most practical
4. When cluster shapes are complex/non-spherical, DBSCAN or spectral may be worth the computational cost