Title: Analyzing the Efficacy of Principal Component Analysis in League of Legends Team Ranking with a Special Note on LPL Teams Data Accuracy

Introduction: In the dynamic and complex world of eSports, assessing the performance and ranking of League of Legends (LoL) teams involves a nuanced understanding of multiple variables. This report elaborates on a Python script that employs Principal Component Analysis (PCA) and explained variance ratios to analyze and rank LoL teams, taking into account data spanning three years.

Methodology: The script is designed to process and analyze datasets containing detailed statistics of LoL teams and players from 2021 to 2023. This chronological data is instrumental in offering insights into the teams' performance progression. Weighting is applied to prioritize recent data, ensuring the rankings reflect the current standing of the teams. The code is meticulous in cleaning the data, extracting only those rows with non-null 'team\_name', thereby ensuring data integrity.

Principal Component Analysis (PCA): PCA is a dimensionality reduction technique instrumental in our methodology. Given the multifaceted nature of the datasets involving variables like building destroyed, champion kill, ward killed, etc., PCA simplifies the complexity by transforming the original variables into a new set of uncorrelated variables, called principal components. These components retain most of the original data’s variance in descending order.

In this script, PCA is applied to twelve selected features. Missing values in these features are imputed using the mean strategy to maintain dataset completeness. The data is then normalized to bring all variables to a comparable scale, crucial for the effectiveness of PCA. The PCA is configured to extract two principal components, reducing the dataset's dimensionality while retaining the essence of the original data.

Special Note on LPL Teams: It is essential to highlight that some data from the LPL (LoL Pro League) is lacking, leading to a potential accuracy issue in the ranking of these specific teams. The absence of complete datasets for LPL teams may affect the comprehensiveness and reliability of their rankings and composite scores derived from the PCA.

Explained Variance Ratios and Application: The explained variance ratios are pivotal in understanding the significance of each principal component extracted through PCA. Teams are then ranked based on a composite score, offering a concise yet comprehensive ranking system considering multiple variables and their evolution over three years.

The script handles datasets from 2021 to 2023, applying a weighting system to underscore the importance of recent performance. The player data is then aggregated by team, factoring in these weights to compute weighted kills, victims, and wards placed and killed counts.

Conclusion: This methodology provides a nuanced, dynamic, and robust approach to ranking LoL teams, combining statistical and machine learning techniques in eSports analytics. However, users should interpret the rankings of LPL teams with caution, given the incomplete data, to ensure a balanced and informed perspective on their performance standings.