The provided code performs a credit risk quantization analysis by categorizing FICO scores into discrete credit rating buckets. The primary objective is to optimally segment the continuous range of FICO scores into a fixed number of buckets in a way that captures meaningful differences in default risk across segments. The method used for this segmentation is based on minimizing the mean squared error (MSE) or maximizing the likelihood, depending on the selected configuration.

The code begins by importing the necessary packages and loading FICO score and default data from a CSV file into a pandas DataFrame. It then groups the data by unique FICO scores to compute the count of borrowers and the number of defaults associated with each score. This grouping simplifies subsequent computations by aggregating identical score entries.

To efficiently compute the segmentation cost for various bucket ranges, the code precomputes cumulative statistics—namely, the cumulative number of observations and defaults. Using these, it calculates a cost matrix that stores the loss (either MSE or negative log-likelihood) for every possible segment between score indices. For MSE, the loss reflects the error incurred by approximating all values in a segment with a single default probability. For likelihood, it reflects how probable the observed defaults are under a binomial distribution.

Dynamic programming is then used to compute the optimal segmentation that minimizes total cost across all buckets. The algorithm constructs a table (dp) to track the minimum cumulative cost of dividing the scores into a given number of buckets. It also tracks backpointers (prev) to reconstruct the optimal cut points that separate the buckets.

After the optimal cuts are identified, the code extracts the corresponding score boundaries and assigns a credit rating label to each borrower based on their score's position within the segmented intervals. Ratings are assigned in descending order—i.e., lower labels correspond to higher FICO scores (better credit).

This quantization approach is particularly useful for converting continuous credit scores into categorical risk ratings, making the data more suitable for downstream modeling tasks that require discrete inputs. It also helps in summarizing and visualizing the risk distribution across a loan portfolio in a more interpretable format.