

## Phase 2 — PCA (Dimensionality Reduction)

1. I loaded the cleaned dataset (303 rows, 25 features + target).
2. I separated the features (X) from the target column.
3. Since the data was already scaled, I applied PCA directly.
4. PCA showed that:
  - PC1 explains ~22% of the variance.
  - PC2 explains ~13%.
5. To keep most of the information:
  - 12 components reach 90% variance.
  - 14 components reach 95% variance.
6. I saved the PCA-transformed dataset and created charts:
  - Bar chart of explained variance ratio.
  - Cumulative variance plot with 90% and 95% lines.
  - Scatter plot of PC1 vs PC2 colored by target.
  - Table of top components with variance values.

### Key Numbers

- Samples: 303
- Features: 25
- PC1: 22% variance
- PC2: 13% variance
- 90% variance → 12 PCs
- 95% variance → 14 PCs

### Visualizations

Explained Variance Ratio (first 20 components)



