Synthetic Clinical Dataset Report — Early-stage Ovarian Cancer

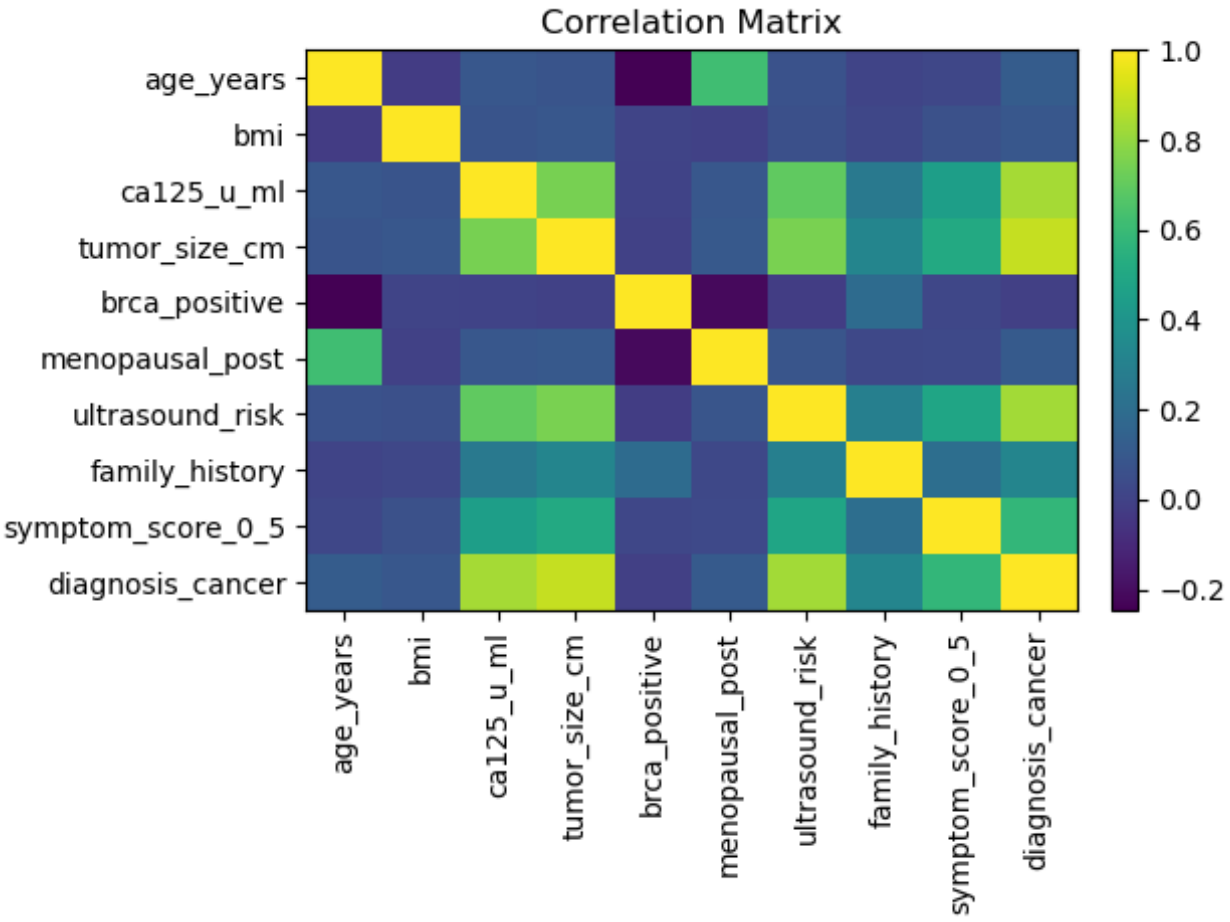
Armenian/Eastern European–tuned parameters

Synthetic Clinical Dataset Report — Early-stage Ovarian Cancer (Armenian/Eastern European–tuned)
Date: 2025-08-13 14:03 UTC

Records: 800
Cancer prevalence: 36.1%
BRCA+ overall prevalence: 14.0%
Healthy with CA-125 < 35 U/mL: 90.8%
Cancer with CA-125 > 200 U/mL: 88.2%

Key Modeling Choices:
- BRCA+ skew younger by 6–8 years; earlier menopause center near 48–49.
- CA-125: log-normal by diagnosis + linear dependence on tumor size; targets enforced for clinical plausibility.
- Ultrasound risk: beta-distributions; small boost with tumor size.
- Family history: higher baseline with BRCA+/cancer reflecting clustering in smaller populations.
- BMI: slightly lower means than US references.

Notes:
- Literature from multiple sources (including international and US-based studies) defined baseline ranges.
  Parameters were then adjusted to reflect Armenian/Eastern European populations and underrepresented BRCA+ subtypes.
- Synthetic data for ML experimentation only; not real patient data.



Note: Synthetic, for research/ML only.