**1. Asset Universe & Clustering**

* **Select ~10–20 liquid crypto perpetual futures (BTC, ETH, SOL, ADA, AVAX, etc.) from major exchanges (Binance, Bybit, FTX, etc.)**
* **Use PCA and dynamic clustering on a broad feature set — including price, order book imbalance, funding rate volatility, realized volatility — to identify groups of coins with strong latent co-movements**
* **Goal: exploit mean reversion within these clusters dynamically, not just static pairs**

Key features evaluated:

* Half-life of mean reversion (how quickly spreads revert)
* Stationarity of spread (ADF test p-value) — only stable spreads qualify
* Spread volatility (rolling std dev) — informs risk & sizing
* Volatility persistence (GARCH alpha + beta) — regime detection
* Average hourly volume — ensures sufficient liquidity
* Funding rate and its volatility — to understand profit/funding dynamics

**2. Pair/Group Selection & Dynamic Cointegration**

* **Test intra-cluster pairs or small groups for:**
  + Cointegration (Engle-Granger / Johansen tests) using rolling windows or Kalman filters to detect regime shifts
  + Stationarity of spreads (ADF test)
  + Half-life of mean reversion
* **Retain only dynamically stable pairs/groups where cointegration persists adaptively**

**3. Feature Engineering & ML Model for Mean Reversion Prediction**

* **Construct a rich feature set including:**
  + Spread z-score, velocity, and acceleration (momentum)
  + Volatility metrics (rolling std, GARCH parameters, ARCH test p-values)
  + Market microstructure indicators (order book imbalance, bid-ask spread, volume spikes)
  + Funding rate and its volatility
  + Regime signals (from step 4)
* **Train an ML model (e.g., LSTM, Random Forest, or Gradient Boosting) to predict:**
  + Probability of spread mean reversion within N bars (classification + probabilistic output)
* **Model explicitly incorporates regime-awareness**

**4. Regime Detection & Strategy Adaptation**

* **Use clustering or Hidden Markov Models on volatility, returns, and funding rate data to identify regimes:**
  + Low vs high volatility
  + Trending vs mean-reverting markets
* **Use regime signals to dynamically:**
  + Adjust trading thresholds
  + Inform ML model inputs
  + Enable/disable trading or vary position sizing

**5. Trading Logic & Execution**

* **Open long/short positions when:**
  + Spread divergence exceeds a dynamic threshold AND
  + ML model predicts high probability of mean reversion
* **Position sizing based on:**
  + Signal confidence
  + Real-time volatility targeting (exposure scaled inversely with realized volatility)
  + Liquidity considerations (bid-ask spread, volume)
* **Implement stop loss and profit taking rules based on spread movement or time limits**
* **Simulate realistic trading costs: slippage, fees, latency**
* **Optional: add market-making elements with adaptive inventory control quoting around predicted fair spreads**

**6. Backtesting & Performance Analysis**

* **Backtest on high-frequency data (1-min or tick level if possible) with realistic cost assumptions**
* **Evaluate key metrics:**
  + Sharpe and Sortino ratios
  + Maximum drawdown
  + Hit rate and trade-level PnL distribution
  + Rolling performance by regime (to verify robustness)
  + ML model diagnostics (confusion matrix, ROC-AUC)
* **Stress test using Monte Carlo simulations across varying volatility and regime scenarios**

**7. Optional GUI / Dashboard**

* **Interactive interface to:**
  + Select asset pairs/clusters and adjust strategy parameters
  + Visualize live & historical spreads, signals, trades, and PnL
  + Toggle ML and regime filters on/off for experimentation
  + Monitor real-time risk and execution metrics