**Project Idea Structure**

**1. Asset Universe Selection**

* Choose ~10 liquid assets (crypto: BTC, ETH, SOL... or equities: ETFs, tech stocks)
* Use PCA / clustering to group related ones (e.g. BTC–ETH, QQQ–XLK)

Features for K-Means Clustering:

* Mean Return Function (90 Days) Function
* Return Volatility Function
* Skewness of returns Function
* Autocorrelation of returns Function
* Augmented Dickey-Fuller (ADF) test p-value Function
* ARCH test p-value Function
* Average daily trading volume Function
* Volume volatility Function
* Max drawdown (rolling over period) Function
* Average RSI (Relative Strength Index) Function
* Average Bollinger Bands Width Function

**2. Pair Selection + Cointegration**

* Test all pairs for:
  + Cointegration (Engle-Granger / Johansen)
  + Stationarity of spread (ADF test)
  + Half-life of mean reversion
* Keep only the most promising pairs

**3. ML Model to Predict Reversions**

* Use features like:
  + Spread Z-score, spread velocity
  + RSI, Bollinger bands of spread
  + Volatility of each asset
  + Add ARCH test p-values or GARCH-predicted volatility as features to your classifier:
  + Helps ML model learn if spread reversion is more/less likely during high-volatility periods.
  + Supports regime-awareness directly in feature space
* Predict:
  + Probability of spread reverting in N bars
  + Classification (1 if spread reverts to mean within time window)

**4. Regime Detection**

* Use clustering or HMM to detect:
  + Low-vol vs high-vol regimes
  + Trending vs mean-reverting periods
* Allow ML model or threshold logic to adapt to regime
* Add ARCH test p-values or GARCH-predicted volatility as features to your classifier:
* Helps ML model learn if spread reversion is more/less likely during high-volatility periods.
* Supports regime-awareness directly in feature space

**5. Trading Logic**

* Long/short pair when:
  + Spread diverges past threshold *and* ML model predicts mean reversion
  + Position sizing based on confidence or volatility targeting

**6. Backtesting & Diagnostics**

* Metrics: Sharpe, Sortino, hit rate, drawdown, PnL curve
* Confusion matrix for ML model
* Rolling performance (e.g. strategy works well only in certain regimes?)

**7. Optional GUI**

* Select pair + strategy config
* Visualise trades, spread, PnL
* Toggle ML or regime-based filters on/off