Here’s a **crisp and well-structured meeting pitch** you can use to **frame your discussion** about Advanced Analytics capabilities for time series and market data quality:

**🧠 Pitch: Advanced Analytics for Market Data – Strategic Framing**

**"In principle, yes — we fully agree that we need a dedicated capability in Advanced Analytics focused on data quality, remediation, and model-aligned time series. But to move from agreement to execution, we need clarity on scope, ownership, and operating model."**

**🎯 Framing Questions:**

1. **Team Structure**

➤ Are we building a new, dedicated Advanced Analytics team under Data?  
➤ Or are we transferring SME knowledge from RA (Risk Analytics) to Data Strats, desk by desk?

1. **Remediation Ownership**

➤ Will the Data team just **implement** fixes based on RA/MRO requirements?  
➤ Or do we also own **remediation decisions** (e.g., proxy choice, smoothing logic)?

1. **Time Series Nuances Vary by Desk**  
   ▪ **FNMA DUS**: Data appears stale in 2008–2009 — but the Desk **agrees** with these marks (gov sources aligned).  
   ▪ **ABS Proxy Contraction**: The fix is to **create new time series** — this can be built in-house or sourced, but **desk prioritization is on CMO**, not ABS today.  
   ▪ **Equity Vol Params**:
   * Operational burden: 5K tickers.
   * Daily calibration needed.
   * Model is subpar.
   * **Outlier smoothing via interpolation is not accepted** — would mask real market moves.

**🧩 Dependencies & Realities**

* Fixes aren’t just technical — they depend on:
  + Desk buy-in
  + RA guidance
  + Prioritization per asset class
  + Model inputs and calibration cycles
* **The data issues are real**, but solutions are:
  + **Desk-specific**
  + **Model-aware**
  + **Capacity-dependent**

**✅ What We Can Own (with Proper Support)**

* **Build new time series** (where gaps are known)
* **Prototype ML-based suspecting / imputation**
* **Pre-onboarding tests for quality + reasonability**
* **Partner with desks to gradually reduce RA dependency**

**⚠️ But to materialize this:**

* We need **dedicated headcount**
* Clear **RACI** for ownership, decision-making, and remediation
* Cross-team alignment on expectations (RA, MRO, Desk)

Would you like this formatted into a **PowerPoint slide** or **summary note** for leadership?

Great question. Missing data imputation techniques are methods used to estimate and fill in missing values in datasets. In quantitative modeling, especially in financial time series like in RFDM or Vasara, choosing the right technique depends on the data structure, business logic, volatility, and statistical assumptions.

Here’s a **comprehensive list** of imputation techniques, grouped by category:

**🟢 1. Simple (Deterministic) Methods**

These are basic methods, easy to implement, often used when data is stable or gaps are small.

| **Technique** | **Description** |
| --- | --- |
| **Forward Fill (Last Value Carried Forward)** | Fill with the last observed value before the missing point. |
| **Backward Fill** | Fill with the next known value after the gap. |
| **Mean/Median Imputation** | Fill with average or median of available data. Not suitable for volatile time series. |
| **Constant Value** | Use a fixed value (e.g., zero, or a domain-specific default). |

**🟡 2. Interpolation-Based Methods**

Great for continuous, trending data — but can **over/understate risk** during volatile periods.

| **Technique** | **Description** |
| --- | --- |
| **Linear Interpolation** | Draws a straight line between known points. Best for short gaps. |
| **Polynomial Interpolation** | Uses a curve (e.g., quadratic) to interpolate. Riskier if overfit. |
| **Spline Interpolation** | Smooth interpolation using piecewise polynomials. Less jagged than linear. |
| **Cubic Hermite / Akima** | Provides more controlled curvature, used in financial curves. |

**🟠 3. Statistical / Probabilistic Methods**

These techniques assume distributions and rely on statistical models.

| **Technique** | **Description** |
| --- | --- |
| **Brownian Bridge** | Assumes a stochastic process to fill between two points. Common for rates or prices. |
| **Expectation Maximization (EM)** | Iterative algorithm estimating missing data by maximizing likelihood. |
| **Kalman Filtering** | Recursive filter modeling time series dynamics; works well with noise. |
| **ARIMA-based Imputation** | Fits time series models to forecast missing values. |

**🔵 4. Machine Learning / AI-Based Techniques**

Useful when data patterns are complex or multivariate.

| **Technique** | **Description** |
| --- | --- |
| **KNN Imputation** | Fills using values from similar (neighboring) time series. |
| **Regression Imputation** | Predict missing value using other variables/features. |
| **Random Forests / XGBoost** | Predictive modeling for missing data. |
| **Neural Networks (RNN, LSTM)** | Learn patterns over time to fill gaps, especially in long sequences. |
| **Autoencoders / GANs** | Advanced models to reconstruct missing segments from latent representations. |

**🔴 5. Domain-Specific / Rule-Based**

Important for financial modeling and risk.

| **Technique** | **Description** |
| --- | --- |
| **Proxy Time Series** | Replace missing series with highly correlated series. |
| **Cohort Mapping** | Use data from similar instruments (e.g., same rating, region, sector). |
| **Floor/Ceiling Rules** | Business-defined thresholds. E.g., "yields must be non-negative." |
| **Rating Hierarchies** | Fill using data from higher or lower credit tiers. |

**⚠️ Things to Consider Before Choosing**

* **Volatility** of the asset class (e.g., linear fill might understate risk in HY credit).
* **Regulatory Use** (e.g., FRTB modellability thresholds).
* **Purpose**: risk model vs valuation vs reporting.
* **Auditability** and explainability.

Would you like me to make this into:

* a slide?
* a PDF reference table?
* or a table specific to **Credit and SPG time series imputation** in RFDM?