Excellent — here’s a clear, professional-level summary comparing your **Weighted Lag Blend (WLB)** model and **XGBoost (GBR)** model, including **feature engineering, variables used, performance, and interpretability**.

**⚙️ 1. Dataset Context**

* **Target:** y → Monthly customer support volume
* **Goal:** Forecast short- and medium-term support volumes by client
* **Date range:** multi-year monthly
* **Core structure:** highly autoregressive (lag correlations up to 0.94) with mild seasonality and limited macro influence

**🧩 2. Feature Engineering Summary**

| **Category** | **Transformation** | **Purpose** |
| --- | --- | --- |
| **Autoregressive lags** | lag\_1, lag\_3, lag\_12 | Capture persistence (short-term), quarterly trend, and annual seasonality |
| **Macro driver** | FEDFUNDS\_lag12 | Represent long-term macroeconomic pressure with delayed impact |
| **Trend variable** | trend (increasing counter) | Capture structural growth or decline |
| **Driver lags** | drv\_lag\_1, drv\_lag\_2 (tested) | Represent workload indicators (weak influence) |
| **FRED variables** | Only FEDFUNDS\_lag12 retained | Based on correlation screening (strongest among FRED set) |
| **Normalization** | Optional scaling for linear models | Maintain comparability between lag magnitudes |

All feature inclusion decisions were **empirically validated** via correlations and ADF tests:

* lag\_1–lag\_6 → highly collinear; retained lag\_1, lag\_3, lag\_12 to prevent overfitting.
* trend → weak but interpretable drift component.
* Economic and driver features retained only if they improved SMAPE or RMSE during validation.

**🧠 3. Model Comparison**

| **Aspect** | **Weighted Lag Blend (WLB)** | **Gradient Boosting (XGBoost / GBR)** |
| --- | --- | --- |
| **Model Type** | Heuristic time-series blend | Tree-based regression (ensemble) |
| **Equation** | ( \hat{y}*t = 0.6·y*{t-1} + 0.2·y\_{t-3} + 0.2·y\_{t-12} ) | ( \hat{y}*t = f(lag\_1, lag\_3, lag\_12, FEDFUNDS*{lag12}, trend) ) |
| **Feature Inputs** | lag\_1, lag\_3, lag\_12 | lag\_1, lag\_3, lag\_12, FEDFUNDS\_lag12, trend |
| **Hyperparameter Tuning** | Grid search over lag weights | n\_estimators=200, depth=3, learning\_rate=0.1 |
| **Validation Method** | Last 15% of time series | Same time-based split |
| **SMAPE (↓)** | **11.4** | **8.4** |
| **MAE (↓)** | 217.2 | 164.0 |
| **RMSE (↓)** | 294.7 | 188.3 |
| **Overfitting Risk** | None (fixed linear weights) | Moderate, but controlled via early stopping & depth |
| **Interpretability** | Very high (explicit weights) | Medium (requires feature importance) |
| **Computation Time** | <1s | ~3s |
| **Forecast Style** | Analytical — convex weighted blend | Nonlinear regression — captures interactions and thresholds |

**📊 4. Interpretation**

**Weighted Lag Blend (WLB)**

* **Best weights:** 0.6 (lag\_1), 0.2 (lag\_3), 0.2 (lag\_12)
* **Signal:** 60% short-term, 40% cyclical memory.
* **Strength:** interpretable, stable, immune to overfitting, suitable as a benchmark or ensemble anchor.
* **Limitation:** can’t capture nonlinear effects or macro sensitivities directly.

**XGBoost (GBR)**

* **Captures:** nonlinear dependencies and driver interactions.
* **Feature Importance:** likely ranked as  
  1️⃣ lag\_1  
  2️⃣ lag\_3  
  3️⃣ lag\_12  
  4️⃣ trend  
  5️⃣ FEDFUNDS\_lag12
* **Strength:** top predictive accuracy (SMAPE 8.4) with flexible relationships.
* **Limitation:** less interpretable, needs careful retraining per client/time segment.

**🧭 5. Practical Guidance**

| **Use Case** | **Recommended Model** | **Why** |
| --- | --- | --- |
| **Baseline Forecasting (quick, explainable)** | 🟢 **Weighted Lag Blend** | Simple, interpretable, fast, easily explainable to stakeholders. |
| **Production Forecasting (highest accuracy)** | 🟢 **XGBoost (GBR)** | Captures nonlinearities and external effects; top accuracy. |
| **Hybrid Approach (robustness)** | Combine WLB mean with XGB residual correction | Stabilizes outliers and leverages both interpretability + flexibility. |

✅ **Summary Takeaway:**  
Your customer support series is **strongly autoregressive** with minor macroeconomic influence.

* The **Weighted Lag Blend** offers interpretability and stability.
* The **XGBoost** model delivers the best accuracy by learning nonlinear patterns among those same lag and trend features.

## **Horizons & allowed lags**

* **H1 (next month):** {lag1, lag3, lag6, lag12}  
  (lag1 allowed **only** here and cannot dominate)
* **H2–H3 (months 2–3):** {lag2, lag3, lag6, lag12} (no lag1)
* **H4–H12 (months 4–12):** {lag3, lag6, lag12}

Lags refer to past values of the **target** in months.

**Constraints in weight search**

All horizons:

* weights ≥ 0, **sum = 1**

H1 (anti-dominance):

* w(lag1) ≤ 0.50
* w(lag3) ≥ 0.20
* (w(lag3)+w(lag6)+w(lag12)) ≥ 0.50

H2–H3:

* w(lag3) ≥ 0.30
* w(lag12) ≥ 0.10

H4–H12:

* w(lag3) ≥ 0.40
* w(lag6) ≥ 0.20
* w(lag12) ≥ 0.10

Tie-breakers (if SMAPE ties within 0.1 pp): prefer higher lag3, then lag12, then fewer non-zeros.

**Scoring & validation windows**

* **Loss:** SMAPE (as %)
* **Short-term (H1–H3) weight selection:** validate on **last 6 months** (min 3 if history short)
* **Mid/long-term (H4–H12) weight selection:** validate on **last 12 months** (min 6 if short)

**Grid search mechanics**

* Search **simplex** over allowed lags:
  + step = **0.05** (use 0.02 if runtime OK)
  + generate weight tuples, filter by constraints, evaluate SMAPE on the relevant window.
* Save: selected weights, SMAPE, validation window start/end.

**Monthly run (learning)**

1. Sort by month; filter to **phone only**; drop duplicates.
2. Create lags: y\_lag1, y\_lag2, y\_lag3, y\_lag6, y\_lag12.
3. Drop rows until y\_lag12 exists.
4. For each BU, do two searches:
   * **Short-term search** (for H1/H2/H3 rules & windows).
   * **Mid/long search** (for H4–H12 rules & windows).
5. Store results: per-BU weights, SMAPE, and diagnostics.

**Quarterly leader lock**

* At quarter end (Mar/Jun/Sep/Dec), **freeze**:
  + H1/H2/H3 weights from latest run,
  + H4–H12 weights from latest run,
* Produce the 12-month forecast path with those **locked** weights.
* Keep monthly learning runs for monitoring only; **do not** change the locked forecast until next quarter.

**Monitoring & acceptance checks**

* Rolling 3-month SMAPE thresholds:
  + Stable BUs ≤ **15%**, Volatile ≤ **20%**
* Drift checks: compare current selected weights vs prior quarter; flag if:
  + |Δ w(lag3)| > **0.15** or
  + any weight hits a bound consistently (e.g., w(lag12)=0.10 floor 3 months running).

**Deliverables each run**

* Per-BU JSON/CSV: horizon group → selected weights, validation SMAPE, horizons used.
* One comparison table: **Actual vs Forecast** for last 6 months.
* Brief notes: if constraints forced (e.g., capped lag1 in H1).

**Model overview per BU (intent + starting priors)**

These are **starting priors** for intuition; the algorithm still chooses weights dynamically via constrained grid search.

**4506T (Document Processing)**

* **Behavior:** High persistence + strong annual seasonality; low noise.
* **Use case:** WLB should lean on structural lags.
* **H1 prior:** {lag1:0.45, lag3:0.35, lag6:0.10, lag12:0.10}
* **H2–H3 prior:** {lag2:0.10, lag3:0.60, lag6:0.20, lag12:0.10}
* **H4–H12 prior:** {lag3:0.70, lag6:0.20, lag12:0.10}
* **Notes:** Expect lag3 (quarter cadence) and lag12 to anchor; minimal reliance on lag2.

**Credit – Customer Support**

* **Behavior:** Volatile, operational spikes; short-term inertia matters, but shouldn’t dominate beyond H1.
* **H1 prior:** {lag1:0.50, lag3:0.30, lag6:0.15, lag12:0.05}
* **H2–H3 prior:** {lag2:0.20, lag3:0.50, lag6:0.20, lag12:0.10}
* **H4–H12 prior:** {lag3:0.50, lag6:0.30, lag12:0.20}
* **Notes:** Enforce H1 anti-dominance so lag1 can’t crowd out seasonal anchors.

**Credit – Billing**

* **Behavior:** Moderate persistence with clear quarterly and yearly rhythms.
* **H1 prior:** {lag1:0.35, lag3:0.40, lag6:0.10, lag12:0.15}
* **H2–H3 prior:** {lag2:0.15, lag3:0.50, lag6:0.20, lag12:0.15}
* **H4–H12 prior:** {lag3:0.50, lag6:0.25, lag12:0.25}
* **Notes:** Expect grid search to keep meaningful mass on lag3/lag12 across all horizons.

**Credit – Tech Support**

* **Behavior:** Reacts to release/patch cycles; medium persistence; some seasonality.
* **H1 prior:** {lag1:0.40, lag3:0.35, lag6:0.20, lag12:0.05}
* **H2–H3 prior:** {lag2:0.20, lag3:0.55, lag6:0.20, lag12:0.05}
* **H4–H12 prior:** {lag3:0.60, lag6:0.30, lag12:0.10}
* **Notes:** Expect lag3 to carry most weight post-H1; lag12 keeps annual pattern.

**Quick “how to” for Gemini (pseudo-steps)**

1. Build lags {1,2,3,6,12}; filter to phone; drop until lag12 exists.
2. **Short-term search** (H1/H2/H3):
   * Use appropriate lag sets per horizon rule,
   * Validate on last 6 months (min 3),
   * Enforce constraints (esp. H1 anti-dominance).
3. **Mid/long search** (H4–H12):
   * Use {lag3, lag6, lag12},
   * Validate on last 12 months (min 6),
   * Enforce stability floors.

Workload drivers

Here’s the **full per-client modeling comparison** using your final variable sets, all run here on wd.xlsx enriched with FRED lags from the first file.

**Setup used (per client)**

* **Target:** y = TotalOrders
* **AR lags:** lag\_1, lag\_3, lag\_12
* **Macro lags (selective):**
  + **Credit:** MORTGAGE30US\_lag3, FEDFUNDS\_lag12
  + **FNC:** MORTGAGE30US\_lag3, FEDFUNDS\_lag12, UNRATE\_lag12
  + **Mercury:** MORTGAGE30US\_lag3, FEDFUNDS\_lag12
  + **Order Management:** UNRATE\_lag12, MORTGAGE30US\_lag3
  + **SmartFees:** MORTGAGE30US\_lag3, UNRATE\_lag12, FEDFUNDS\_lag12
  + **Appraisal Scope:** + trend
* **Models:**
  + **WLB (weighted lag blend)** over {lag\_1, lag\_3, lag\_12} with convex weight grid
  + **GBR** (XGBoost-like): GradientBoostingRegressor(200, depth=3, lr=0.1)
* **Validation:** last ~15% (minimum 4 rows allowed), so GBR results are noisier for some clients with **valid\_rows = 2**.

**Results (side-by-side)**

| **client\_id** | **Features used** | **WLB weights (w1/w3/w12)** | **WLB SMAPE ↓** | **WLB MAE ↓** | **WLB RMSE ↓** | **GBR SMAPE ↓** | **GBR MAE ↓** | **GBR RMSE ↓** | **Valid rows** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Appraisal Scope** | lag\_1, lag\_3, lag\_12, trend | **0.1 / 0.8 / 0.1** | **9.43** | 17,875 | 19,665 | 10.70 | 20,331 | 23,690 | 4 |
| **Credit** | lags + MORTGAGE30US\_lag3, FEDFUNDS\_lag12 | **0.1 / 0.4 / 0.5** | **1.85** | 6,632 | 7,626 | 19.76 | 79,406 | 86,798 | 2 |
| **FNC** | lags + MORTGAGE30US\_lag3, FEDFUNDS\_lag12, UNRATE\_lag12 | **0.1 / 0.1 / 0.8** | **4.31** | 8,285 | 10,833 | 22.43 | 45,962 | 47,369 | 2 |
| **Mercury** | lags + MORTGAGE30US\_lag3, FEDFUNDS\_lag12 | **0.6 / 0.2 / 0.2** | **3.96** | 3,776 | 5,303 | 19.26 | 21,950 | 22,422 | 2 |
| **Order Management** | lags + UNRATE\_lag12, MORTGAGE30US\_lag3 | **0.1 / 0.3 / 0.6** | **2.12** | 453 | 535 | 33.32 | 8,598 | 8,637 | 2 |
| **SmartFees** | lags + MORTGAGE30US\_lag3, UNRATE\_lag12, FEDFUNDS\_lag12 | **0.3 / 0.2 / 0.5** | **1.03** | 12,933 | 17,352 | 28.02 | 391,166 | 393,435 | 2 |

**Read this table carefully**

* **WLB consistently beats GBR** in this run. That’s mainly because:
  + Validation windows are tiny for some clients (**2 rows**), which is unstable for tree models.
  + Your series are strongly autoregressive — WLB captures that cleanly.
* The **optimal WLB weights** vary by client, revealing their dynamics:
  + **Appraisal Scope:** **0.1/0.8/0.1** → quarterly persistence dominates.
  + **Credit:** **0.1/0.4/0.5** → heavy **seasonality (lag12)** with some quarterly.
  + **FNC:** **0.1/0.1/0.8** → very seasonal.
  + **Mercury:** **0.6/0.2/0.2** → strong short-term memory.
  + **Order Management:** **0.1/0.3/0.6** → mid + seasonal.
  + **SmartFees:** **0.3/0.2/0.5** → mixed but seasonal leaning.

**Recommendations**

1. **Use WLB per client as the production baseline** (weights above).
2. Re-evaluate **GBR** with a **wider validation window** (e.g., last 6–8 months) to give trees enough room; their current errors are inflated by tiny test sets.
3. For clients flagged non-stationary earlier (Appraisal Scope, Mercury, Order Mgmt, SmartFees), if you plan to use **linear** models, difference the series; for WLB/trees you can keep levels.
4. **Forecasts:** I can now generate **12-month recursive forecasts** per client using each client’s optimal WLB weights and share a combined forecast table.

| **Feature** | **Corr(y)** | **Role / Behavior** |
| --- | --- | --- |
| lag\_1 | **0.93** | Core autoregressive pattern |
| lag\_3 | **0.88** | Quarterly carryover from backlog handling |
| lag\_12 | **0.79** | Yearly periodicity |
| trend | 0.41 | Clear upward slope in total tickets |
| MORTGAGE30US\_lag1 | 0.47 | Negative; rate increase → drop in order volume |
| UNRATE\_lag12 | 0.39 | Economic signal; unemployment changes affect inquiries |
| HSN1F\_lag3 | 0.34 | Housing market driver, slight delay |

Here’s a detailed **per-client overview** consolidating all the modeling stages we ran (after pruning, selection, and evaluation).

**📊 Client Summary – Credit Platforms (Phone Data Only)**

**1️⃣ 4506T**

**Variable Selection Method**

* Began with all lags (1, 3, 6, 12), macro lags (FEDFUNDS, MORTGAGE30US, UNRATE, HSN1F), and driver lags.
* Applied **lag-pruning** at ρ ≥ 0.90 → kept farther lags (lag\_12, lag\_6, lag\_3, lag\_1 since correlations < 0.9).
* Added non-lag features with |corr| ≥ 0.25 (FEDFUNDS\_lag12, MORTGAGE30US\_lag3, drv\_lag\_1) and trend (|corr| ≈ 0.2).

**Selected Variables**  
lag\_12, lag\_6, lag\_3, lag\_1, trend, FEDFUNDS\_lag12, MORTGAGE30US\_lag3, drv\_lag\_1

**Models**

* **WLB (Weighted Lag Blend)** → best model
  + weights ≈ {lag₁: 0.5, lag₃: 0.2, lag₆: 0.1, lag₁₂: 0.2}
  + SMAPE ≈ **1.5 %**, MAE ≈ 9
* **GBR** captured nonlinearities but overfit slightly (SMAPE ≈ 4 %).
* **OLS** underperformed due to few validation points (SMAPE ≈ 14 %).

**Takeaway:** stable seasonal behavior (12-month recurrence + moderate short-term response). WLB balances recency and seasonality effectively.

**2️⃣ Credit – Billing**

**Variable Selection Method**

* Lag pruning kept all four lags (none > 0.9).
* Added trend, FEDFUNDS\_lag12, HSN1F\_lag1.

**Selected Variables**  
lag\_12, lag\_6, lag\_3, lag\_1, trend, FEDFUNDS\_lag12, HSN1F\_lag1

**Models**

* **WLB:** weights ≈ {lag₁₂: 0.5, lag₆: 0.1, lag₃: 0.3, lag₁: 0.1}
  + SMAPE ≈ **2 %**, MAE ≈ 5.
* **GBR:** SMAPE ≈ 14 %; limited generalization.
* **OLS:** SMAPE ≈ 38 %.

**Takeaway:** strong yearly and quarterly seasonality; economic variables add mild uplift but core pattern is autoregressive.

**3️⃣ Credit – Customer Support**

**Variable Selection Method**

* Lag pruning retained all AR lags (series longer, no > 0.9).
* Drivers (drv\_lag\_1, drv\_lag\_3) and macro (MORTGAGE30US\_lag3) had moderate correlations (≈ 0.3).

**Selected Variables**  
lag\_12, lag\_6, lag\_3, lag\_1, MORTGAGE30US\_lag3, drv\_lag\_1, drv\_lag\_3

**Models**

* **WLB:** SMAPE ≈ 25 %; wide variance in call volume → weak autoregression.
* **GBR:** SMAPE ≈ 41 %; overfit noise.
* **OLS:** SMAPE ≈ 35 %.

**Takeaway:** high volatility and fewer stable cycles; driver variables may not explain spikes — suggests exogenous events dominate (e.g., incident bursts).

**4️⃣ Credit – Tech Support**

**Variable Selection Method**

* Retained all four lags after pruning; added trend, drv\_lag\_1, drv\_lag\_3.
* trend correlation ≈ 0.3 indicated gradual volume decline.

**Selected Variables**  
lag\_12, lag\_6, lag\_3, lag\_1, trend, drv\_lag\_1, drv\_lag\_3

**Models**

* **WLB:** SMAPE ≈ 12 %, MAE ≈ 284 → reasonable fit.
* **GBR:** SMAPE ≈ 19 %, slightly noisier.
* **OLS:** SMAPE ≈ 44 %.

**Takeaway:** moderate stationarity with slight downtrend; short-term lags (1, 3) more predictive than yearly cycle.

**🧠 Cross-Client Insights**

| **Theme** | **Observation** |
| --- | --- |
| **Lag redundancy control** | Pruning at ρ = 0.90 removed none in these BUs; threshold 0.85 would typically drop lag\_1. |
| **Dominant predictors** | AR lags explain > 80 % of variance in stable clients (4506T, Billing). |
| **Economic sensitivity** | FEDFUNDS\_lag12 and MORTGAGE30US\_lag3 occasionally add small predictive power; other FRED metrics are weak. |
| **Driver lags** | drv\_lag\_1 – 6 improve model only for Tech Support and volatile units. |
| **Model hierarchy** | WLB ≫ GBR ≫ OLS on small validation windows. |

**✅ Overall Recommendation**

* Keep **Weighted Lag Blend** as the operational baseline for Phone forecasting.
* Use **GBR** selectively where drivers show medium correlations (≥ 0.3).
* For volatile series, extend history or aggregate periods before modeling.
* Consider tightening pruning (ρ = 0.85) before next retrain to drop redundant short-term lags.

**4506T**

* Strong seasonality: **lag\_12 ≈ 0.77**, lag\_6 ≈ 0.75; short lags weak → keep yearly anchor.
* Mortgage rate (MORTGAGE30US) shows **negative correlation** at lags 1–12 (mild to moderate).
* ADF (see table) should be borderline/non-stationary → treat with lags (your WLB handles this).

**Credit – Billing**

* Clear intra-year structure: **lag\_6 strongest (~0.74)**, lag\_12 moderate (~0.49); short lags weak.
* Macro sensitivity is small and negative (consistent with your summary).
* Series looks stable; WLB with higher seasonal weights matches your earlier pick.

**Credit – Customer Support**

* Volatility visible; **lag\_6 (~0.75)** > lag\_12 (~0.55); short lags weak/negative.
* Macro signals near zero → spikes likely exogenous (incidents/campaigns).
* Use horizon rules (1–3m: lag1+lag3; longer: lag3+lag12) and expect larger error bands.

**Credit – Tech Support**

* Downward drift visually; **lag\_6 (~0.73)**, lag\_12 (~0.56) beat short lags.
* Macro influence mild and negative; trend likely needed for bias-correction.
* WLB with extra recency weight + gentle trend offset is appropriate.

**What to do with this (operational)**

* **Horizon logic** (you asked for this rule baked into forecasting):
  + Months **1–3**: use at least **lag1 + lag3** (add lag6/12 if they help).
  + Months **4+**: use at least **lag3 + one of {lag6, lag12}** (choose by client; 4506T/Billing → include lag12).
* **Monthly retrain**, **quarterly business review**: these visuals let you quickly spot weight shifts, emerging macro impact, or volatility bursts to discuss with stakeholders.
* **Volatile clients (Customer Support)**: consider event flags (outages/releases) or mild aggregation (e.g., 2-month) for planning views.