Seller Prioritisation Playbook v3.0

MSc Dissertation Framework

Author: Oladapo Obadina
Institution: Robert Gordon University
Date: January 2026
Version: 3.0 (Revised with 210-day threshold)

Abstract

This playbook presents a comprehensive, academically rigorous framework for predicting long-term seller value in online marketplaces, with a focus on operational prioritisation under resource constraints. Built around a decision-focused machine learning philosophy and the CRISP-DM methodology, the framework ensures models drive concrete resource allocation decisions rather than just predictions.

The methodology is validated through an MSc dissertation using the Olist Brazilian e-commerce dataset, with explicit testing of whether features beyond early revenue improve prioritisation quality.

Key updates in v3.0:

- **210-day observation window** (was 365 days) - enables larger sample size and temporal validation
- **150-day future window** for label construction (was 305 days)
- **2,043 eligible sellers** (was 1,170) - 74% increase in sample size
- **Temporal holdout option** - can test on 2018 sellers
- All methodological improvements from v2.0 retained

Table of Contents

---

1. Business Understanding (CRISP-DM Phase 1)

1.1 Background and Problem Context

Online marketplaces depend on third-party sellers for revenue, assortment, and growth. However, seller performance is highly concentrated: empirical analysis of the Olist dataset reveals that the top 25% of sellers generate approximately 85% of gross merchandise value (GMV). This concentration creates a resource allocation challenge for marketplace operations teams.

The Core Problem: RevOps teams have limited capacity (typically can actively support 15-25% of sellers). They must decide which new sellers warrant investment in onboarding support, account management attention, and promotional incentives. Without data-driven prioritisation, resources are spread inefficiently.

1.2 Business Objectives

Primary Objective:
Enable marketplace operations teams to identify high-potential sellers early in their lifecycle (within the first 60 days of activity) so that resources can be prioritised toward sellers most likely to drive future GMV.

Secondary Objectives:

1. Understand which early-window behaviours and characteristics predict long-term seller success
2. Determine whether predictive models using rich behavioural features add meaningful value over simple early-revenue ranking
3. Provide interpretable insights that translate to actionable interventions

1.3 Success Criteria

Success is defined across four dimensions:

Technical Discrimination:

Criterion	Threshold	Rationale
Test AUC-ROC	> Baseline	Must beat early GMV only model
Test AUC-PR	> Baseline	Important for imbalanced classes

Business Impact:

Criterion	Threshold	Rationale
GMV Capture at Top-20%	> Baseline	Must beat baseline performance
Lift over Random	> 2.5x	Meaningful improvement over no model

Statistical Rigour:

Criterion	Threshold	Rationale
Significance level	$p < 0.05$	Results not due to chance
Effect size	Report Cohen's d	Practical significance

Interpretability:

Criterion	Threshold	Rationale
Named Actionable Insights	>= 3	Must provide practical guidance
Feature Importance Stability	> 0.7 correlation across folds	Robust, not spurious findings

Minimum Viable Outcome:
If the full model does not meaningfully beat the baseline, the project still delivers value by:

1. Confirming that simple heuristics suffice for seller prioritisation
2. Surfacing insights about seller behaviour via SHAP analysis
3. Providing a rigorous methodological blueprint for other marketplace contexts

1.4 Situation Assessment

Available Resources:

Resource	Description
Data	Olist Brazilian E-Commerce dataset (public, Kaggle)
Computing	Google Colaboratory / local Python environment
Time	8 weeks (7 working + 1 buffer)
Expertise	MSc Business Analytics candidate

Constraints:

Constraint	Description	Mitigation
Single marketplace	Brazil only, 2016-2018	Frame as methodological contribution
No off-platform data	Cannot see external marketing, seller characteristics	Acknowledge in limitations
Observational only	No A/B testing possible	Focus on prediction, not causation
Public dataset	Results reproducible but not novel data	Contribution is framework, not data

Key Assumptions:

1. Early-window behaviour (days 0-59) contains predictive signal about medium-term performance
2. The GMV concentration curve provides a natural threshold for defining high-value
3. Patterns in Olist 2016-2018 are indicative of general marketplace dynamics
4. Delivered orders are the appropriate basis for measuring seller success
5. First order date is the appropriate anchor for seller tenure
6. **150-day future window (~5 months) is sufficient to identify seller success patterns**

1.5 Research Questions

RQ1: To what extent can early-window seller behaviour (first 60 days) predict medium-term seller value (GMV in days 60-210)?

RQ2: Does a model incorporating behavioural, operational, and product mix features outperform a baseline using early GMV alone?

RQ3: Which early-window features are most predictive of seller success, and what do they reveal about seller ramp-up dynamics?

RQ4: What predicts seller success when early revenue signals are excluded, and how does this differ from the full-feature model?

1.6 Risk Assessment

Risk	Likelihood	Impact	Mitigation
Early GMV dominates completely	Medium	Medium	Frame as valid finding; Model B tests GMV-independent signals
Insufficient sample size	Low	High	210-day threshold provides 2,043 sellers
Class imbalance causes majority prediction	Medium	Medium	Class weighting, threshold tuning, AUC-PR evaluation

Risk	Likelihood	Impact	Mitigation
Temporal patterns do not generalise	Medium	Medium	Temporal holdout validation available with 210-day threshold
Findings not generalisable beyond Olist	Medium	Low	Frame contribution as methodology; acknowledge in limitations
Feature leakage	Low	High	Strict as-of rules; EDA on train set only
150-day window misses slow-ramp sellers	Medium	Medium	Sensitivity analysis comparing to longer windows

1.7 Project Plan (CRISP-DM Mapped)

Week	CRISP-DM Phase	Key Activities
1	Phase 2-3	Data loading, exploration, eligibility filtering, label construction, baseline model
2	Phase 3-4	Data splitting, EDA on train set, feature engineering, feature selection
3	Phase 4	Model A (all features), Model B (excluding early GMV), cross-validation
4	Phase 4-5	Hyperparameter tuning, final evaluation, SHAP analysis
5	Phase 5	Robustness checks, statistical significance, sensitivity analysis
6	Phase 5-6	Literature review, methodology chapter
7	Phase 6	Discussion, conclusion, dissertation drafting
8	Buffer	Supervisor feedback, final polish

---

2. Data Understanding (CRISP-DM Phase 2)

2.1 Data Source

Dataset: Olist Brazilian E-Commerce Public Dataset
Source: Kaggle (https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce)
Period: September 2016 - October 2018
Geography: Brazil

2.2 Data Schema

Table	Rows	Key Columns	Role in Analysis
orders	99,441	order_id, customer_id, order_purchase_timestamp, order_status	Transaction timestamps, status filtering
order_items	112,650	order_id, seller_id, product_id, price, freight_value	Links sellers to orders, GMV calculation
sellers	3,095	seller_id, seller_state, seller_city	Unit of analysis, metadata
products	32,951	product_id, product_category_name	Product mix features
customers	99,441	customer_id, customer_unique_id	Customer behaviour features
reviews	99,224	order_id, review_score	Quality signals
payments	103,886	order_id, payment_value	Alternative GMV calculation
geolocation	1,000,163	zip_code, lat, lng	Not used (quality issues)

2.3 Data Quality Assessment

Quality Checks Performed:

Check	Result	Action
Duplicate order_ids	None found	No action
Missing seller_ids in order_items	None	No action
Orders with null timestamps	0	No action
Cancelled/pending orders	~1%	Exclude from GMV calculations
Reviews without orders	217	Exclude (orphan records)
Geolocation inconsistencies	Significant	Do not use geolocation table

Olist-Specific Considerations:

1. **Data ends October 2018:** Patterns may not reflect current marketplace dynamics
2. **Brazilian market only:** Currency (BRL), logistics, consumer behaviour are Brazil-specific
3. **Unknown marketplace interventions:** Olist may have already prioritised some sellers, creating selection bias
4. **No seller demographics:** Cannot control for seller experience, resources, or business type

2.4 Observation Window Trade-off Analysis

Critical Decision: Why 210 days instead of 365 days?

Threshold	Future Window	Total Sellers	2018 Sellers	Temporal Holdout?
365 days	305 days (~10 months)	1,392	0	❌ Impossible
270 days	210 days (~7 months)	1,797	82	⚠️ Marginal
210 days	**150 days (~5 months)**	**2,043**	**328**	✅ **Good**
180 days	120 days (~4 months)	2,213	498	✅ Good
150 days	90 days (~3 months)	2,406	691	✅ Good

Selected: 210-day threshold

Rationale:

1. **74% more sellers** (2,043 vs 1,170) - significantly more statistical power
2. **Samples-per-feature ratio: 84** (vs 48) - much safer from overfitting
3. **Enables temporal holdout** - can test on 328 sellers from 2018
4. **150-day future window is meaningful** - ~5 months captures medium-term success patterns
5. **Balances label quality with sample size** - not too short (90 days) or impractical (365 days with no 2018 data)

Limitation acknowledged: May miss slow-ramp sellers who take 6+ months to reach peak. Sensitivity analysis will compare results across different thresholds.

---

3. Data Preparation (CRISP-DM Phase 3)

3.1 Unit and Temporal Definitions

Element	Definition	Rationale
Unit of Analysis	seller_id	Business decision is per-seller
Anchor Date	Date of seller's first delivered order	Marks start of seller lifecycle
Early Window	Days 0-59 after anchor	Features computed here only
Future Window	**Days 60-210 after anchor**	**Label computed here (150 days)**

Element	Definition	Rationale
Observation Requirement	**210+ days after anchor**	**Ensures complete label window**

3.2 Data Splitting Strategy

CRITICAL: Data must be split BEFORE any exploratory analysis.

Split Strategy:

```
Total eligible sellers: 2,043

Option A: Random Split (70/15/15)
├── Train: 1,430 sellers (70%) - EDA, feature selection, cross-validation
├── Validation: 306 sellers (15%) - Hyperparameter tuning
└── Test: 307 sellers (15%) - Final evaluation (touched ONCE)

Option B: Temporal Holdout
├── Train + Val: ~1,715 sellers (Pre-2018) - All model development
└── Test: ~328 sellers (2018) - Final evaluation (mimics real deployment)
```

Recommended: Option A (Random Split) for primary analysis, with **Option B (Temporal Holdout)** as robustness check.

3.3 As-Of Rule Enforcement

Golden Rule: Every feature must be computable using only data available by end of day 59 after first order.

```
FOR EACH seller:
  anchor_date = first_order_date
  feature_cutoff = anchor_date + 59 days
  label_start = anchor_date + 60 days
  label_end = anchor_date + 210 days

  Features use data WHERE event_date <= feature_cutoff
  Labels use data WHERE label_start <= event_date <= label_end
```

3.4 Explicit Leakage Prevention

Prohibited:

1. Any metric computed from day 60-210 data used as feature
2. Future order dates or counts
3. EDA or feature selection using test/validation data
4. Cohort-level statistics including target seller's future data

Verification Steps:

1. Data validation gates (8 checks)
2. Temporal leakage verification (5 checks)
3. All feature names prefixed with "early_"

3.5 Eligibility Criteria

Criterion	Threshold	Rationale
Observable days	>= 210	Complete label window
Delivered orders	>= 1 in early window	Active seller
Extreme outliers	< 99.9th percentile early GMV	Remove anomalies

Sample Size:

- Total sellers with delivered orders: 2,970
- After 210-day filter: 2,043
- After outlier removal: ~2,040

3.6 Label Construction

Target Variable:

```
future_gmv = SUM(price)
   WHERE order_date BETWEEN anchor_date + 60 AND anchor_date + 210
   AND order_status = 'delivered'

high_value = 1 IF future_gmv >= threshold (top 25%)
           0 OTHERWISE
```

Threshold Selection:

- Use GMV concentration curve on training set
- Primary threshold: Top 25% captures ~85% of GMV
- Test robustness across 20%, 25%, 30%

Expected Class Distribution:

- High-value (1): ~25% (~510 sellers)
- Non-high-value (0): ~75% (~1,533 sellers)
- Class ratio: 1:3 (manageable imbalance)

3.7 Data Validation Gates

All Must Pass Before Modeling:

Check	Validation
seller_id uniqueness	No duplicates in modeling table
anchor_date completeness	Not null for all sellers
Feature window integrity	No feature events after day 59
Label window integrity	No label events before day 60 or after day 210
Temporal consistency	delivery_date >= order_date
Value sanity	GMV >= 0, quantities >= 0
Label distribution	20-30% high-value (not extreme imbalance)
No test data in EDA	Verify split before any analysis

---

4. Modeling (CRISP-DM Phase 4)

4.1 Feature Engineering

4.1.1 Feature Design Principles

1. **Temporal Integrity:** Strict day 0-59 window enforcement
2. **EDA on Train Set Only:** Feature selection based on training data patterns
3. **Reproducibility:** Same logic in training and production
4. **Stability:** Rates and ratios preferred over raw counts
5. **Interpretability:** Clear business meaning for each feature

4.1.2 EDA-Informed Feature Selection

Process:

1. Calculate feature correlations with target (train set only)
2. Calculate feature-feature correlations (train set only)
3. Drop features with |correlation with target| < 0.05 (weak predictors)
4. Drop one feature from pairs with |correlation| > 0.85 (redundancy)

Results from EDA:

- Original features: 40
- After removing weak predictors: ~35
- After removing redundant features: 17
- Samples-per-feature ratio: 84.1 (excellent)

4.1.3 Final Feature Set (17 features)

Model A Features:

#	Feature	Category	Hypothesis
1	early_gmv	Volume	Higher early revenue predicts success
2	early_order_count	Volume	More orders = more traction
3	early_freight_avg	Volume	Per-item shipping cost
4	early_aov	Volume	Higher AOV = premium positioning
5	early_freight_ratio	Volume	Shipping burden indicator
6	early_active_weeks	Trajectory	Consistency matters
7	early_last_week_gmv	Trajectory	Recent momentum
8	early_avg_weekly_growth	Trajectory	Growth trajectory
9	early_unique_products	Product Mix	Catalog breadth
10	early_unique_categories	Product Mix	Diversification
11	early_category_hhi	Product Mix	Specialist vs generalist
12	early_avg_review_score	Quality	Quality signal
13	early_review_score_std	Quality	Consistency
14	early_pct_1_star	Quality	Problem rate
15	early_review_trend	Quality	Quality trajectory
16	early_max_delivery_delay	Operations	Worst case reliability
17	early_anchor_quarter	Metadata	Seasonality

Model B Features (14): Same as above, excluding:

- early_gmv

- early_aov
- early_last_week_gmv

4.2 Class Imbalance Handling

Class Distribution: ~25% high-value, ~75% non-high-value (ratio 1:3)

Primary Strategy: Class Weighting

- `class_weight='balanced'` in Logistic Regression
- `scale_pos_weight=3.0` in XGBoost

Why not SMOTE?

- Tested empirically: class weighting outperformed SMOTE
- Class weighting adjusts loss function without modifying data distribution
- Simpler and more robust

4.3 Model Selection

4.3.1 Algorithms (As Committed to Supervisor)

Model	Purpose	Rationale
Logistic Regression	Interpretability anchor	Linear, stable, provides coefficients
Decision Tree	Transparent segmentation	Visual rules, explainable
XGBoost	Performance benchmark	Captures non-linearities

4.3.2 Model Specifications

Logistic Regression:

```python
LogisticRegression(
    random_state=42,
    max_iter=1000,
    class_weight='balanced',
    C=1.0,  # Default regularisation
    penalty='l2'
)
```

Decision Tree:

```python
```

```
DecisionTreeClassifier(
    max_depth=5,
    min_samples_leaf=20,
    class_weight='balanced',
    random_state=42
)
```

XGBoost:

```python
XGBClassifier(
    n_estimators=100,
    max_depth=5,
    learning_rate=0.1,
    min_child_weight=5,
    subsample=0.8,
    colsample_bytree=0.8,
    scale_pos_weight=3.0,
    eval_metric='logloss',
    random_state=42
)
```

4.3.3 Training-Validation Strategy

5-Fold Stratified Cross-Validation:

- Performed on training set only
- Provides robust performance estimates (mean ± std)
- Used for model selection

Hyperparameter Tuning:

- Grid search on validation set
- Parameters tested: C (LR), max_depth (trees)
- Select based on validation AUC

Final Model:

- Train on Train + Validation combined
- Evaluate once on Test set

4.4 Baseline Model

Baseline: Early GMV Only

- Single feature: early_gmv
- Logistic Regression with class_weight='balanced'
- Represents "simple heuristic" that practitioners would use

---

5. Evaluation (CRISP-DM Phase 5)

5.1 Primary Evaluation Metrics

5.1.1 Discrimination Metrics

Metric	Purpose
AUC-ROC	Overall ranking ability
AUC-PR	Performance on positive class (better for imbalance)

5.1.2 Decision-Focused Metrics

Metric	Formula	Interpretation
Precision@K%	TP in top K% / K% of sellers	Quality of prioritisation list
Recall@K%	TP in top K% / Total positives	Coverage of high-value sellers
GMV Capture@K%	GMV of top K% / Total GMV	Business value captured
Lift@K%	GMV Capture@K% / K%	Improvement over random

Primary Metric: GMV Capture@20% with Lift over Baseline

5.2 Statistical Significance Testing

5.2.1 Bootstrap Procedure

```python
FOR i = 1 TO 10,000:
    Sample test sellers with replacement (n = test set size)
    Calculate metric for Model A
    Calculate metric for Baseline
    Store difference

Calculate 95% confidence interval
Calculate p-value (proportion of differences <= 0)
```

5.2.2 Effect Size

Measure	Formula	Thresholds
Cohen's d	(mean_A - mean_B) / pooled_std	Small: 0.2, Medium: 0.5, Large: 0.8

5.2.3 Multiple Comparison Correction

When testing multiple hypotheses:

- Model A vs Baseline
- Model B vs Baseline
- Model A vs Model B

Apply Bonferroni correction: α_adjusted = 0.05 / 3 = 0.017

5.3 Results Reporting Template

Primary Results Table:

Model	Features	Test AUC-ROC	Test AUC-PR	GMV Capture@20%	p-value
Random	-	0.500	0.250	20.0%	-
Baseline	1	TBD	TBD	TBD	-
Model A	17	TBD	TBD	TBD	TBD
Model B	14	TBD	TBD	TBD	TBD

---

6. Deployment (CRISP-DM Phase 6)

6.1 Production Scoring Workflow

Batch Scoring Pipeline:

1. Identify eligible sellers (completed 60 days since first order)
2. Extract features (strictly days 0-59)
3. Apply pre-processing (scaling, imputation)
4. Generate predictions and probabilities
5. Rank sellers by probability
6. Assign priority bands (A/B/C)
7. Generate explanations (top SHAP values)
8. Map to recommended actions
9. Write to output table
10. Alert RevOps team

6.2 Output Schema

Column	Type	Description
seller_id	VARCHAR	Primary key
score_date	DATE	Date of scoring
score	FLOAT	Probability of high-value
priority_band	VARCHAR	A (top 20%), B (20-50%), C (bottom 50%)
top_reason_1	VARCHAR	Primary driver (SHAP)
top_reason_2	VARCHAR	Secondary driver
recommended_action	VARCHAR	Intervention suggestion

6.3 Monitoring Framework

Metric	Frequency	Alert Threshold
Rolling AUC (30-day)	Weekly	Degradation > 0.05
GMV Capture@20%	Monthly	Degradation > 5pp
Feature drift (PSI)	Weekly	PSI > 0.25

---

7. Interpretability and Actionability

7.1 Global Feature Importance

Methods:

1. Logistic Regression coefficients (standardised)
2. SHAP values (model-agnostic)
3. Permutation importance

7.2 Action Mapping Matrix

Pattern	Indicators	Recommended Intervention
High growth + High quality	Active weeks > 4, Review > 4.5	Expansion support
High demand + Fulfillment issues	Orders high, Delays > 5 days	Logistics support
Low freight ratio	freight_ratio < 0.15	Highlight as efficient
Category concentration	HHI > 0.9	Diversification coaching
Declining trajectory	Growth < 0	Re-engagement outreach

7.3 Actionable Insight Format

Template: "Sellers with [feature] [above/below] [threshold] have [X]% probability of becoming high-value, compared to [Y]% for those [below/above] threshold."

---

8. Robustness and Validation

8.1 Sensitivity Analysis

8.1.1 Observation Window Sensitivity

Window	Future Period	Sellers	Test AUC	GMV Capture@20%
180 days	120 days	2,213	TBD	TBD
210 days	**150 days**	**2,043**	**Primary**	**Primary**
270 days	210 days	1,797	TBD	TBD

8.1.2 Threshold Sensitivity

Threshold	High-Value N	Class Ratio	Test AUC
Top 20%	~409	1:4	TBD
Top 25%	~511	1:3	Primary
Top 30%	~613	1:2.3	TBD

8.2 Temporal Robustness

Temporal Holdout (Optional):

Split	Train Period	Test Period	Test AUC
Random	Mixed	Mixed	Primary
Temporal	Pre-2018 (~1,715)	2018 (~328)	Robustness

8.3 Feature Importance Stability

Calculate Spearman correlation of importance rankings across 5 CV folds.

Stability Threshold: Rank correlation > 0.7 across folds

---

9. Ethical Considerations

9.1 Algorithmic Bias Assessment

Source	Risk	Mitigation
Regional bias	Model favours SP (largest state)	Segment analysis
Category bias	Some categories easier to predict	Per-category metrics
Size bias	Larger early sellers favoured	Model B (no early GMV)
Survivorship bias	Only transacting sellers observed	Acknowledge in limitations

9.2 Rich-Get-Richer Risk

Mitigation:

1. Control group: Reserve 10-20% for random selection
2. Acknowledge prediction ≠ causation
3. Regular recalibration

4. Transparency in methodology

9.3 Transparency

- All decisions explainable via SHAP
- No "black box" decisions
- Methodology documented in playbook

---

10. Dissertation-Specific Components

10.1 Academic Contribution

Empirical Contributions:

1. Seller value prediction study on Olist dataset
2. Evidence on whether features beyond early revenue add predictive value
3. Identification of specific early behaviours predicting success
4. Demonstration that GMV features are not essential (Model B finding)

Methodological Contributions:

1. Dual-model approach isolating revenue vs non-revenue signals
2. Decision-focused evaluation framework
3. Rigorous temporal integrity framework
4. Observation window trade-off analysis

10.2 Limitations

Data Limitations:

1. Single marketplace (Olist), single country (Brazil), specific time period (2016-2018)
2. No visibility into off-platform factors
3. No seller demographics
4. 150-day future window may miss slow-ramp sellers

Methodological Limitations:

1. Observational study - cannot establish causation
2. Cannot measure intervention effects
3. Patterns may not generalise to current dynamics

Generalisability:

1. Findings specific to Brazilian e-commerce context
2. Framework is transferable; specific coefficients require re-estimation

10.3 Examiner Defence Preparation

Question	Response
"Why 210 days, not 365?"	Trade-off analysis: 365 days gives 0 sellers from 2018, 210 days gives 2,043 sellers with 328 from 2018, enabling temporal validation
"Why 60 days early window?"	Industry benchmark, balances early intervention with sufficient signal
"How do you know this generalises?"	Framework is transferable; acknowledge coefficients require re-estimation
"Why not causal inference?"	Would require randomised data; frame as predictive study
"What if early GMV dominates?"	Model B tests GMV-independent signals; found they perform nearly as well

---

11. Appendices

Appendix A: Complete Feature Definitions

[17 features with SQL/Python pseudocode]

Appendix B: EDA Outputs

- Univariate distributions
- Bivariate correlations with target
- Feature correlation heatmap

Appendix C: Model Results

- Cross-validation scores
- Test set metrics
- Bootstrap confidence intervals

Appendix D: Code Repository Structure

```
/
├── data/                # Data files (gitignored)
```

```
├── notebooks/
│   ├── 01_data_exploration.ipynb
│   ├── 02_feature_engineering.ipynb
│   ├── 03_modeling.ipynb
│   ├── 04_evaluation.ipynb
│   └── 05_interpretation.ipynb
├── src/
│   ├── data_preparation.py
│   ├── feature_engineering.py
│   ├── modeling.py
│   └── evaluation.py
├── outputs/
├── README.md
├── requirements.txt
└── config.yaml
```

Appendix E: Glossary

Term	Definition
Anchor Date	Date of seller's first delivered order
Early Window	Days 0-59 after anchor date
Future Window	Days 60-210 after anchor date (150 days)
GMV	Gross Merchandise Value
High-Value Seller	Seller in top 25% by future GMV
SHAP	SHapley Additive exPlanations
AUC-ROC	Area Under ROC Curve
AUC-PR	Area Under Precision-Recall Curve

Document Control

Version	Date	Changes
1.0	Jan 2026	Initial playbook
2.0	Jan 2026	CRISP-DM, class imbalance, statistical rigour
3.0	**Jan 2026**	**210-day threshold, 2,043 sellers, temporal holdout option, EDA on train only**

Key Changes from v2.0 to v3.0

Aspect	v2.0 (365-day)	v3.0 (210-day)
Observation window	365 days	210 days
Future window	305 days (~10 mo)	150 days (~5 mo)
Total sellers	1,170	2,043
Train size	819	1,430
Test size	176	307
Samples-per-feature	48	84
2018 sellers available	0	328
Temporal holdout	❌ Impossible	✅ Possible
Overfitting risk	Higher	Lower

End of Playbook v3.0