BREWLYTICS

AI-Powered Offer Completion Prediction with Customer Segmentation
A Technical Presentation for Data Scientists

By: **Airees Rondain**
Post-Graduate Diploma in Artificial Intelligence and Machine Learning
Asian Institute of Management
Makati, Philippines

WHERE CUSTOMER SEGMENTS MEET SMART PROMOTIONS

BREWLYTICS

GOAL

Predict which customers will complete promotional offers to optimize marketing ROI and personalize offer recommendations using customer segmentation.

DATASET OVERVIEW

Dataset	Records	Features	Description
Customer	17,000	5	Demographics (gender, age, income, membership date)
Offers	10	6	Offers characteristics (type, difficulty, duration, channels)
Events	306,534	4	Transaction logs (offers received/viewed/completed, transactions)
Processed Data for Classification	86,432	28	Dataset to be used for classification modeling
Customer Behavior Analysis	16,994	15	Dataset for Customer Behavior Analysis

PROBLEM & DATA OVERVIEW

BREWLYTICS

DATASET CHARACTERISTICS

- **Target Variable:** `offer_completed` (binary: 1=completed, 0=not completed)
- **Class Balance:** 53.4% completed / 46.6% not completed (well-balanced)
- **Time Period:** 2013-2018 customer data
- **Train/Test Split:** 80/20 stratified (69,145 train / 17,287 test)

KEY DATA QUALITY ISSUES IDENTIFIED

- **Missing Demographics:** 12.8% (2,175 customers) missing age/income/gender
- **Age Placeholder:** Age=118 used for missing values
- **MNAR Pattern:** Missingness is Not At Random - customers chose not to provide info

PROBLEM & DATA OVERVIEW

BREWLYTICS

CUSTOMER DEMOGRAPHICS PROFILE

Attribute	Key Insight
Gender	57% Male, 41% Female, 1.4% Other
Age	Peak 50-60 years, multimodal distribution
Income	Multimodal: peaks at $50K. $60K, $75K
Tenure	70% joined 2016-2018 (recent customers)

KEY CORRELATION WITH COMPLETION

Feature	Correlation	Insight
duration	+0.352	Longer offers = higher completion
income	+0.316	Higher income = higher completion
difficulty	+0.270	Counter-intuitive: harder offers correlate with completion
Offer_type_discount	+0.250	Discounts drive completion

OFFER FUNNEL ANALYSIS

Received (76,277) 75.5% → **Viewed (57,725)** 58.1% → **Completed (33,579)** 44.0% overall

- **Biggest Drop-off**: 24.3% never view offers (awareness issue)
- **Secondary Drop-off**: 41.9% view but don't complete (relevance/value issue)

CRITICAL FINDING: MISSING DATA CRISIS

- Customers with missing demographics: **16.8% completion rate**
- Customers with complete demographics: **55% completion rate**
- **3.3x performance gap** - data quality directly impacts outcomes

EDA KEY FINDINGS

BREWLYTICS

PIPELINE OVERVIEW

Raw data (86Kx28) → Drop Leakage (86Kx20) → Encode (86Kx25) → Train/Test Split → Impute → Scale → Ready for ML (24 features)

DATA LEAKAGE PREVENTION (Critical)

Dropped Feature	Reason
‘offer_completed’	Perfect correlation with target (r=1.0)
‘offer_viewed’	Temporal leak - occurs AFTER prediction time
‘completion_time’	Only exists for completed offers
‘time_to_action’	Only exists for completed offers

FINAL FEATURE SET: 24 FEATURES

- 6 binary (channel flags, demographics_missing flag)
- 7 one-hot encoded (3 offer types + 4 gender categories)
- 3 ordinal encoded (age_group, income_bracket, tenure_group)
- 11 scaled numerical features

ENCODING STRATEGY

Type	Features	Method	Rationale
Nominal	offer_type, gender	One-Hot	No inherent order
Ordinal	age_group, income_bracket, tenure_group	Ordinal (0-4)	Natural ordering exists
Binary	channel_flags (email, mobile, social, web)	Unchanged	Alrady 0/1
Numerical	age, income, difficulty, duration, etc.	StandardScaler	Z-score normalization

MISSING VALUE HANDLING

- **Strategy**: Median computation for ‘tenure_group_encoded’
- **Affected**: 87 rows (0.13%) in training set
- **Rationale**: Robust to outliers, preserves ordinal scale, no sample loss

FEATURE ENGINEERING

BREWLYTICS

MODELS EVALUATED

Model	Type	Key Hyperparameters
Logistic Regression	Linear Baseline	max_iter = 1000, class_weight='balanced'
Decision Tree	Tree Baseline	max_depth=10, min_samples_spli=50
Random Forest	Ensemble	n_estimators=100, class_weight='balanced'
XGBoost	Gradient Boosting	n_estimators=100, eval_metrics='logloss'

HYPERPARAMETER TUNING (Random Forest)

- **Method:** GridSearchCV with 5-fold CV
- **Parameter Grid**: 81 combinations tested
    - n_estimators: [50, 100, 200]
    - max_depth: [10, 20, none]
    - min_samples_split: [20, 50, 100]
    - min_samples_leaf: [1, 5, 10]

EVALUATION METHODOLOGY

- **Primary Metric:** F1-Score (balances precision and recall)
- **Secondary Metrics**: AUC-ROC, Accuracy, Precision, Recall
- **Validation**: 5-fold Stratified Cross-Validation
- **Test Holdout**: 20% (17,287 samples)

OVERFITTING ANALYSIS

- **Train-Test Gap:** 15.34% (trainL 99.88%, test:84.54%)
- **CV-test Alignment**: CV F1 (0.8515) vs Test F1 (0.8601) - only 0.9% difference
- **Conclusion**: Model generalizes well despite high training accuracy (typical RF behavior)

MODEL SELECTION & METHODOLOGY

PERFORMANCE COMPARISON (TEST SET)

Model	F1-Score	AUC-ROC	Accuracy	Precision
Random Forest	0.8601	0.9277	0.8454	0.8318
XGBoost	0.8515	0.9154	0.8350	0.8197
Decision Tree	0.8263	0.8984	0.8122	0.8164
Logistic Regression	0.8240	0.8885	0.8056	0.7977

PCA DIMENSIONALITY REDUCTION

Variant	Component	F1-Score	Feature Reduction
Full	24	0.8601	Baseline
90% Variance	8	0.8472	-67% (best trade-off)
95% Variance	10	0.8411	-58%
80% Variance	6	0.8463	-75%

KEY RESULTS

- Best Model: Random Forest
- Performance Lift: +4.4% F1 over Logistic Regression baseline
- CV Stability: All models show std < 0.006 (excellent consistency)
- Recall Focus: 89.05% - captures most offer completions

HYPERPARAMETER TUNING RESULTS

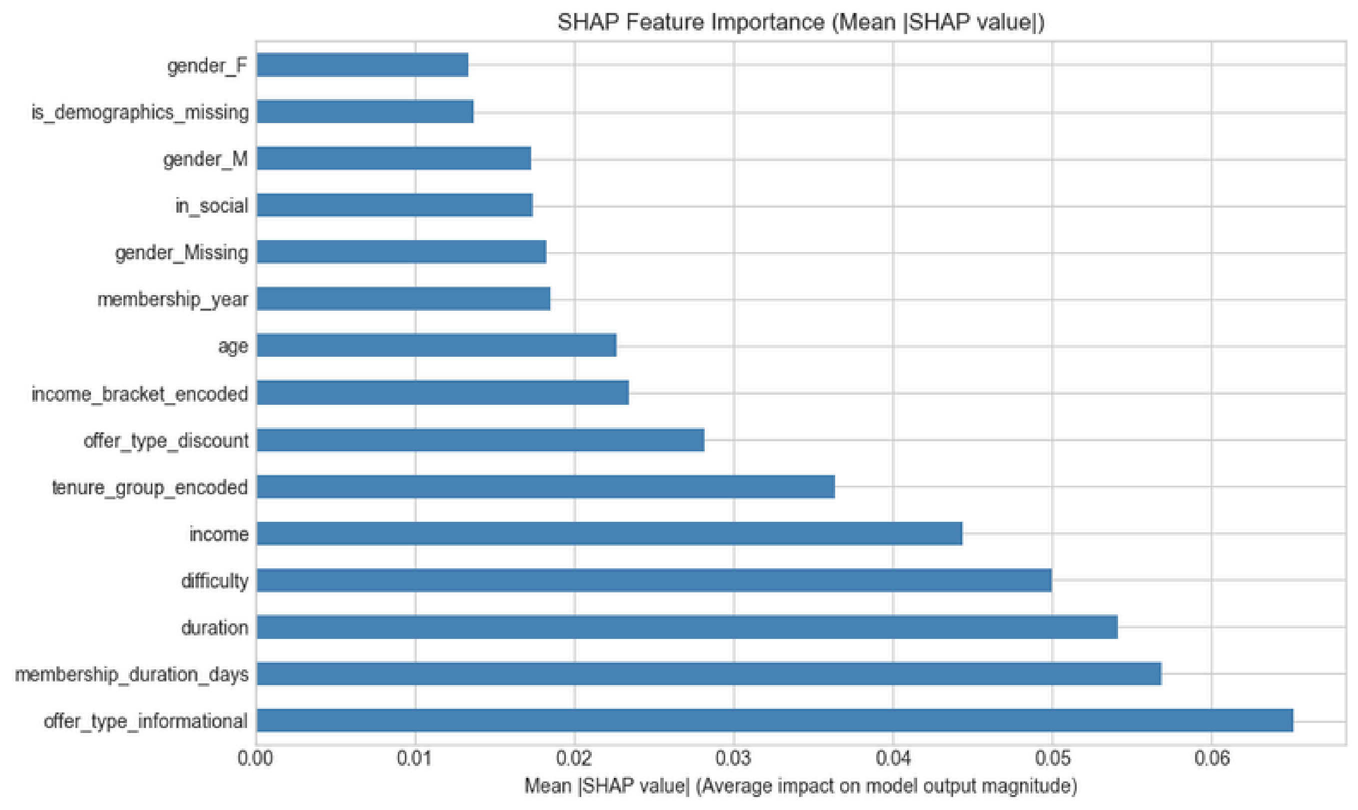
- Best CV F1: 0.8434 (tuned) vs 0.8601 (baseline)
- Result: Tuning decreased performance by 1.60%
- Decision: Use baseline Random Forest (already well-optimized)

RECOMMENDATION

- Use 8-component PCA for production (67% fewer features, only 1.3% F1 drop)

RESULTS & MODEL COMPARISON

GLOBAL FEATURE IMPORTANCE (SHAP Mean Value)



- Top 3 features account for 44.84% of prediction power

BUSINESS INTERPRETATION

- Discount offers are 4x more impactful than BOGO for driving completions
- Shorter, easier offers perform better (contrary to correlation analysis)
- Offer design (52%) > Demographics (34%) - controllable factors dominate

FEATURE CATEGORY IMPORTANCE

Category	% of Total	Insight
Offer Attributes	52.33%	What you offer matters MOST
Demographics	33.55%	Who you target matters
Behavioral	7.67%	Past engagement signals
Channels	4.39%	How you reach them

Key SHAP Insights (Directional Impact)

Feature	Positive SHAP	Negative SHAP	Net Direction
duration	+0.045	-0.191	Longer → Lower completion
difficulty	+0.032	-0.085	Harder → Lower completion
offer_type_discount	+0.085	-0.021	Discount → Higher completion

FEATURE IMPORTANCE & EXPLAINABILITY (SHAP)

BREWLYTICS

K-MEANS CLUSTERING RESULTS (K-5)

Cluster	Name	Size	Completion Rate	Key Profile
0	New Male Members	26,644 (31%)	44.6%	100% M, 0.7 yr tenure, $61K
1	Affluent Female Members	31,279 (36%)	65.9%	100% F, 1.3 yr tenure, $72K
2	Missing Demographics	9,963 (12%)	15.7%	Data quality issue
3	Small Engaged Segment	1,075 (1%)	63.2%	Other gender, 88% view rate
4	Long-Tenure Males	17,471 (20%)	65.3%	94% M, 2.9 yr tenure

CLUSTER QUALITY METRICS

- Silhouette Score: 0.3392 (moderate separation)
- Calinski-Harabasz: 38,554.85 (best among K=3,4,5
- Davies-Bouldin: 0.9918 (lowest = best)

PCA COMPONENT INTERPRETATION

PC	Explained Variance	Top Features
PC1	23.1%	Customer Tenure (membership_year, duration_days)
PC2	14.7%	Demographics (age, income)
PC3	12.3%	Offer Characteristics (difficulty, duration)

CUSTOMER SEGMENTATION

BREWLYTICS

FAIRNESS FRAMEWORK

- Metric: Disparate Impact (80% rule: 0.8 <= DI <=1.25
- Protected Attributes: Gender, Age, Income, Tenure
- Reference Groups: Female, 61-75 age, Very High Income

DISPARATE IMPACT RESULTS

GENDER FAIRNESS

Group	Positive Rate	DI Ratio	Status
Female (ref)	73.48%	1.00	Reference
Male	54.44%	0.741	⚠ UNFAIR
Missing	13.51%	0.184	🔴 CRITICAL
Other	68.52%	0.932	✓ Fair

Intersectional Analysis

- Best: Female × 76+ → 75.52% positive rate, F1=0.9276
- Worst: Male × 18-30 → 40.57% positive rate, F1=0.7645
- Gap: 35 percentage points in positive predictions

Overall Fairness Grade: 🔴 FAILS

- 7 of 18 subgroups fail the 80% rule
- ~58% of test set affected by fairness violations

AGE FAIRNESS

Group	Positive Rate	DI Ratio	Status
61-75 (ref)	66.76%	1.00	Reference
18-30	47.83%	0.717	⚠ UNFAIR
76+	38.18%	0.572	⚠ UNFAIR

INCOME FAIRNESS

Group	Positive Rate	DI Ratio	Status
Very High (ref)	79.62%	1.00	Reference
Low	43.64%	0.548	⚠ UNFAIR
Medium	53.50%	0.672	⚠ UNFAIR
Missing	13.51%	0.170	🔴 CRITICAL

BIAS & FAIRNESS ANALYSIS

LIMITATIONS & TECHNICAL DEBT

DATA LIMITATIONS

1. Missing Data Crisis: 12.8% customers missing demographics
   - Creates systematic 4.3x performance gap
   - Not addressable by modeling alone - requires data collection fix
2. Temporal Coverage: Data only through July 2018
   - Customer behavior may have evolved
   - Model needs periodic retraining
3. Sample Imbalance in Protected Groups:
   - "Other" gender: only 1.2% of dataset
   - Statistical tests may be underpowered for small groups

MODEL LIMITATIONS

1. Overfitting Signature: 15.3% train-test accuracy gap
   - Acceptable given strong CV consistency
   - Monitor in production for drift
2. Feature Leakage Risk:
   - Removed offer_viewed (temporal leak)
   - Some behavioral features may still leak information
3. Static Model:
   - Doesn't account for temporal dynamics
   - No online learning / continuous adaptation

FAIRNESS LIMITATIONS

1. Algorithmic Bias:
   - Model encodes demographic patterns (33% importance)
   - Young males systematically disadvantaged (35% gap)
2. Feedback Loop Risk:
   - Under-predicted groups receive fewer offers
   - Leads to less engagement data → worse predictions → cycle continues

BREWLYTICS

IMMEDIATE ACTIONS (Production Deployment)

Model Recommendation

- Deploy: Baseline Random Forest (F1: 0.8601, AUC: 0.9277)
- Alternative: 8-component PCA model for latency-sensitive applications
  - 67% fewer features, only 1.3% F1 drop

Fairness Mitigation

1. Test demographic-free model: Remove gender, age, income
   - Expected: ~2% accuracy drop, major fairness improvement
2. Threshold calibration: Adjust per-group thresholds to equalize TPR
3. Monitor weekly: Track DI ratios by protected attribute

Data Quality

1. Fix onboarding: Make demographics required fields
2. Incentivize completion: Profile completion campaign (+500 points)
3. Flag in predictions: Separate handling for missing-data customers

FUTURE TECHNICAL WORK
Model Improvements

Initiative	Expected Impact
Time-Series Features	Capture seasonality, trends
Online Learning	Adapt to concept drift
Fairness-aware algorithms	Adversarial debiasing
Causal Inference	Understand true offer effects

RECOMMENDATIONS & FUTURE WORK

SUMMARY

PROJECT ACHIEVEMENTS

- ✅ Built production-ready offer completion prediction model (F1: 0.8601)
- ✅ Identified 24 clean features free of data leakage
- ✅ Achieved 92.77% AUC-ROC for class separation
- ✅ Created 5 actionable customer segments
- ✅ Quantified bias across 4 protected attributes
- ✅ Deployed interactive Streamlit application

THE BOTTOM LINE

Offer design matters more than customer targeting. A well-designed offer (discount, short, easy) outperforms a poorly-designed offer sent to the "perfect" customer segment. Focus optimization on the controllable 52% before demographic targeting.