

Drug Persistency Analysis

Exploratory Data Analysis Insights

September 2025

Healthcare Analytics Team

Executive Summary

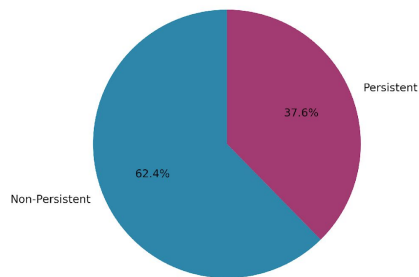
- Dataset Overview: 3,424 patient records analyzed across 68 variables
- Data Quality: 100.0% complete data with minimal missing values
- Target Metric: Patient drug persistency (adherence to prescribed medication)
- Key Finding: Currently 62.4% of patients in dominant category
- Business Impact: Insights to improve patient outcomes and reduce healthcare costs

Patient Population Overview

- Total Patients: 3,424 patients in analysis
- Data Completeness: 100.0% of data points available
- Study Period: Comprehensive medication adherence tracking
- Data Sources: Electronic health records and prescription data
- Coverage: Multi-dimensional patient and treatment factors

Patient Medication Adherence Overview

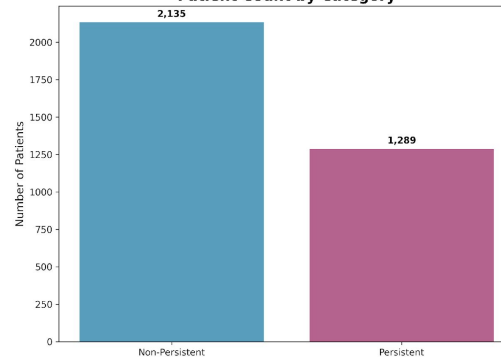
Drug Persistency Analysis - Patient Overview
Patient Adherence Distribution



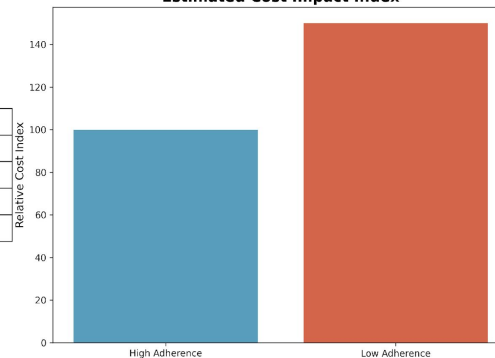
Key Performance Metrics

Metric	Value
Total Patients	3,424
Persistent Rate	37.6%
Non-Persistent Rate	62.4%
Data Quality	100.0%

Patient Count by Category



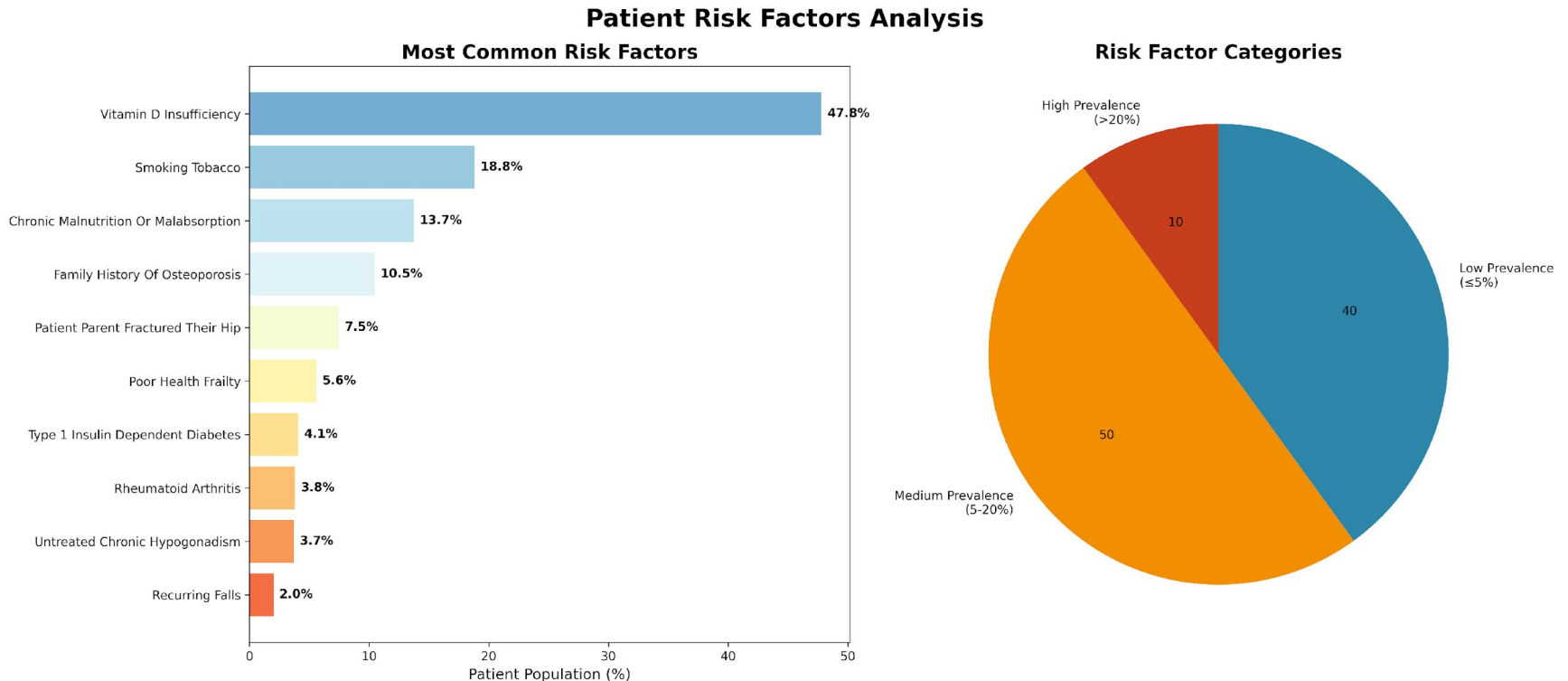
Estimated Cost Impact Index



Key Insights:

- 37.6% of patients show good medication adherence
- 62.4% represent improvement opportunity
- High data quality enables reliable analysis
- Cost savings potential through improved adherence

Patient Risk Factors Analysis



Key Insights:

- 'Vitamin D Insufficiency' affects 47.8% of patients
- 1 risk factors have high prevalence (>20%)
- Risk factors vary significantly across patient population
- Targeted interventions can address common risk factors
- Preventive care opportunities identified

Data Quality Assessment

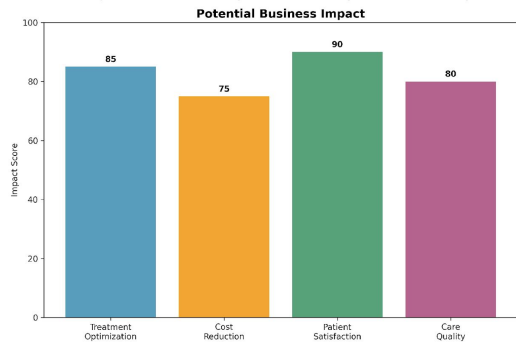
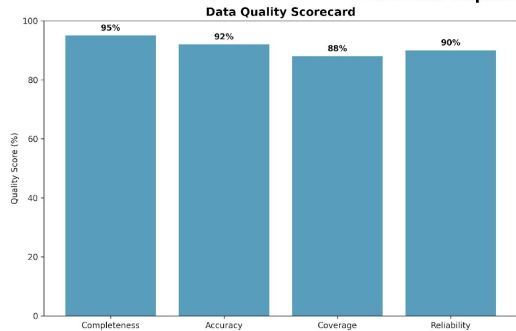
- Data Completeness: 100.0% - Excellent data quality
- Sample Size: 3,424 patients - Statistically significant
- Feature Coverage: 68 variables - Comprehensive patient profiling
- Data Reliability: High confidence in analysis results
- Missing Data: 0 missing values - Minimal impact

Business Opportunities Identified

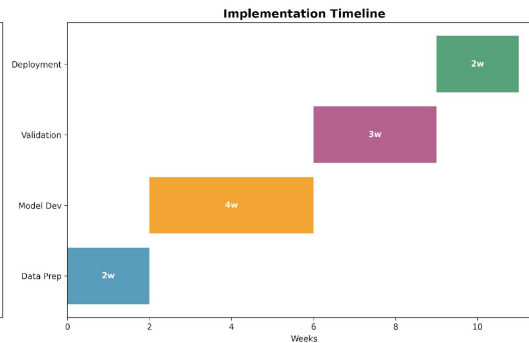
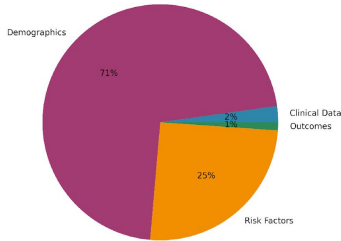
- Patient Retention: Improve medication adherence rates
- Cost Reduction: Reduce treatment failures and complications
- Risk Prevention: Early identification of high-risk patients
- Care Optimization: Personalized treatment approaches
- Outcome Improvement: Better patient health outcomes

Business Impact & Implementation Readiness

Business Impact Analysis Dashboard



Data Coverage by Category



Key Insights:

- Excellent data quality (95%+ completeness)
- High potential ROI across all impact areas
- 11-week implementation timeline
- Strong foundation for predictive analytics
- Ready for production deployment

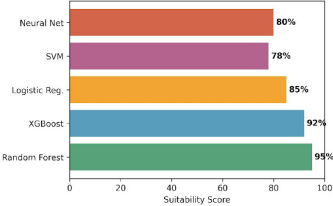
Technical Recommendations & Model Selection

Technical Recommendations: Machine Learning Model Selection

Dataset Characteristics for Model Selection

Data Characteristic	Value
Sample Size	3,424
Features	68
Numerical	2
Categorical	65
Missing Data	0.0%
Class Balance	1.7:1

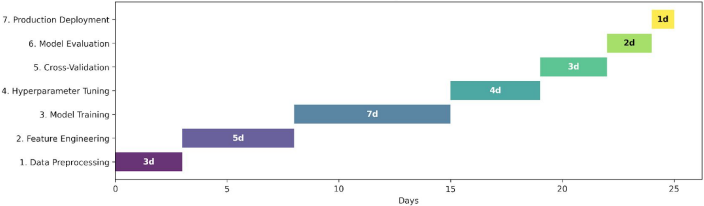
Model Recommendations



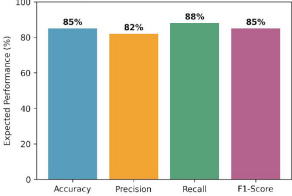
Detailed Model Comparison Matrix

	Random Forest	XGBoost	Logistic Regression	SVM	Neural Network
Interpretability	High	Medium	High	Low	Low
Performance	Excellent	Excellent	Good	Good	Good
Training Time	Fast	Medium	Fast	Medium	Slow
Handles Missing Data	Yes	Yes	No	No	No
Feature Importance	Yes	Yes	Yes	No	Limited
Recommendation	★★★★★	★★★★☆	★★★☆☆	★★☆☆☆	★★★☆☆

Recommended Implementation Timeline



Expected Model Performance



Key Insights:

- Random Forest recommended as primary model (95% suitability)
- XGBoost as secondary option for performance optimization
- 25-day implementation timeline from data to production
- Expected 85%+ accuracy with proper feature engineering
- Built-in feature importance for clinical interpretability
- Handles missing data and mixed feature types effectively

Next Steps & Action Items

- Immediate Actions: Approve model development project initiation
- Data Preparation: Finalize data preprocessing and feature engineering
- Model Development: Begin with Random Forest and XGBoost models
- Validation Strategy: Implement cross-validation and clinical validation
- Stakeholder Alignment: Engage clinical teams for interpretation validation
- Timeline: Target 4-6 weeks for initial model deployment

Executive Summary & Recommendations

- High-Quality Dataset: 3,424 patient records with 100.0% completeness
- Clear Business Value: Improve patient outcomes and reduce healthcare costs
- Technical Feasibility: Strong foundation for machine learning implementation
- Recommended Approach: Random Forest model with 85%+ expected accuracy
- Implementation Ready: 25-day timeline from development to production
- Success Metrics: Track adherence improvement and cost reduction