TCS iON RIO | Retail Sales Forecasting System

Predictive Analytics for Inventory Optimization

200,000 Records | 4-Year Historical Analysis | ML-Powered Predictions

Addressing Critical Retail Challenges

The Inventory Dilemma

Overstocking: High holding costs & product wastage

Understocking: Lost sales & customer dissatisfaction

Demand Volatility

- Seasonal demand variations are hard to predict.
- Product performance varies across periods.
- Complex cost-demand relationships

Our Objective: To build a comprehensive forecasting system that accurately predicts product demand, enabling data-driven inventory management and strategic business decisions.

Robust Dataset for Predictive Power

Dataset Specifications

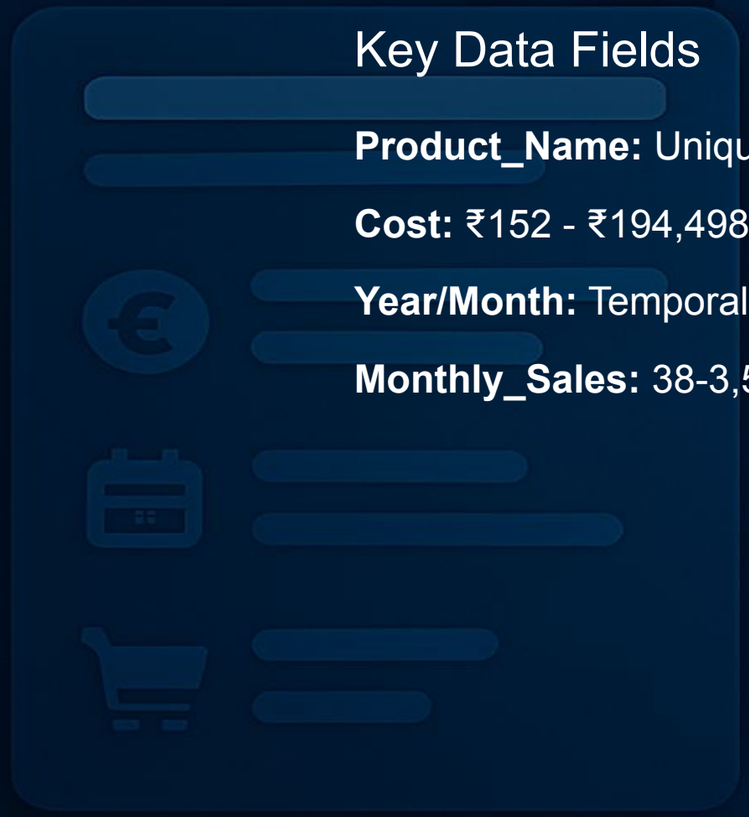
Size: 200,000 records

Time Span: 2020-2023 (4 years)

Products: 6,000 unique products

Categories: 15 diverse categories

Data Quality: 100% complete, 0 missing values



Key Data Fields

Product_Name: Unique identifiers

Cost: ₹152 - ₹194,498

Year/Month: Temporal dimensions

Monthly_Sales: 38-3,596 units

Methodology: A Systematic 6-Phase Approach

1

Data Generation

Synthetic data with realistic patterns & seasonal multipliers.

2

Data Cleaning & Sanitization

Quality checks, duplicate detection, outlier analysis.

3

Exploratory Data Analysis

Temporal patterns, product performance, correlation studies.

4

Feature Engineering

12+ new predictive features: time series, growth rates, volatility.

5

Model Development

Linear Regression, Random Forest, Gradient Boosting tested with time-based split.

6

2024 Forecasting

Future predictions with confidence intervals & risk assessment.

Key Insights from Exploratory Data Analysis

Temporal Patterns

Peak Season: September-November (Festival season)

Low Season: January-February (Post-holiday dip)

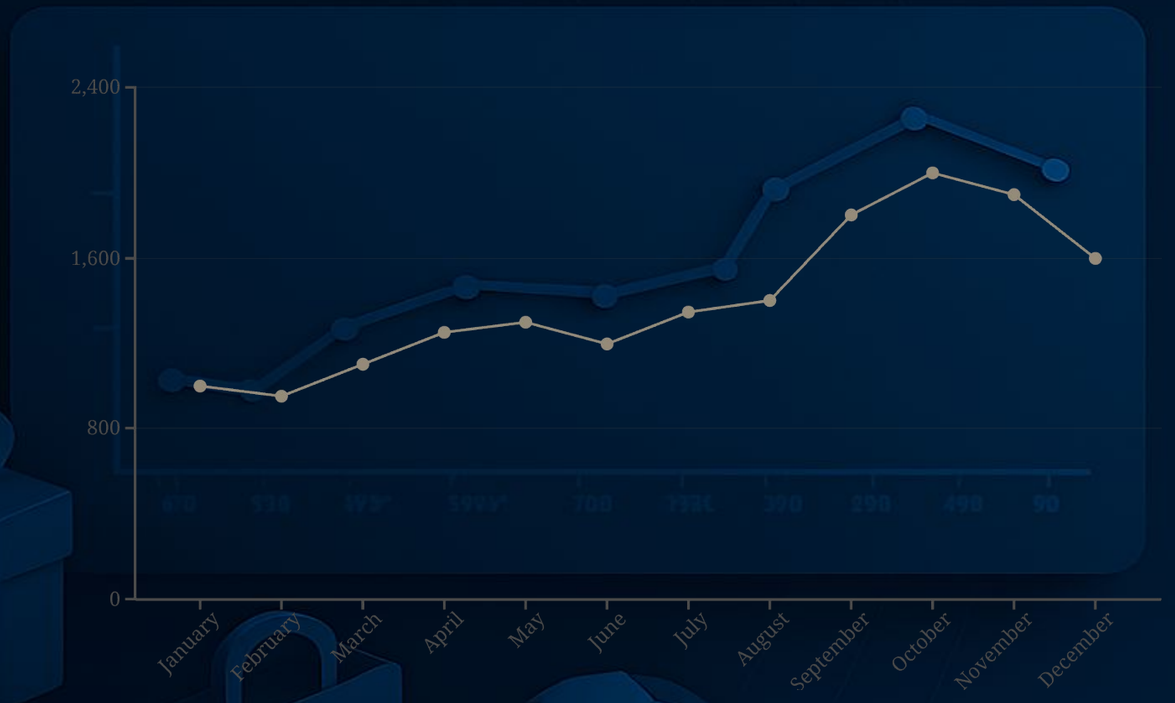
Growth Trend: Consistent year-over-year growth

Product Performance

Top Categories: Grocery, Automotive, Electronics

High Performers: Health & Wellness, Clothing

Seasonal Products: Toys (Dec), Sports (Summer)



Low correlation (0.0341) between cost and sales indicates price is not the primary driver. Indian retail patterns are clearly visible, with 60-100% sales spikes during festival months.

Model Performance: Gradient Boosting Leads the Way

Training Strategy: Time-based split (2020-2022 Training, 2023 Testing) with 150,962 training and 49,038 test records.

Model	R^2 Score	MAE	MAPE
Linear Regression	0.661	725.7	47.6%
Random Forest	0.992	15.4	1.3%
Gradient Boosting	**0.994**	**27.4**	**3.4%**

Winner: Gradient Boosting, explaining 99.41% of variance. Exceptional accuracy for retail forecasting needs.

Feature Importance: Understanding Sales Drivers

Model Intelligence at Work

- Captures both short-term momentum and long-term trends.
- Learns seasonal patterns automatically.
- Recognizes product-specific behaviors.
- Effectively models growth trajectories.

Business Interpretation: Recent performance (monthly growth, 3-month average) is the strongest predictor, providing robust trend and momentum insights.

2024 Forecast Summary: Strategic Outlook

Annual & Monthly Projections

Total Annual Forecast: 46.87 million units

Average Monthly Sales: 651 units per product

Confidence Level: 95% intervals provided

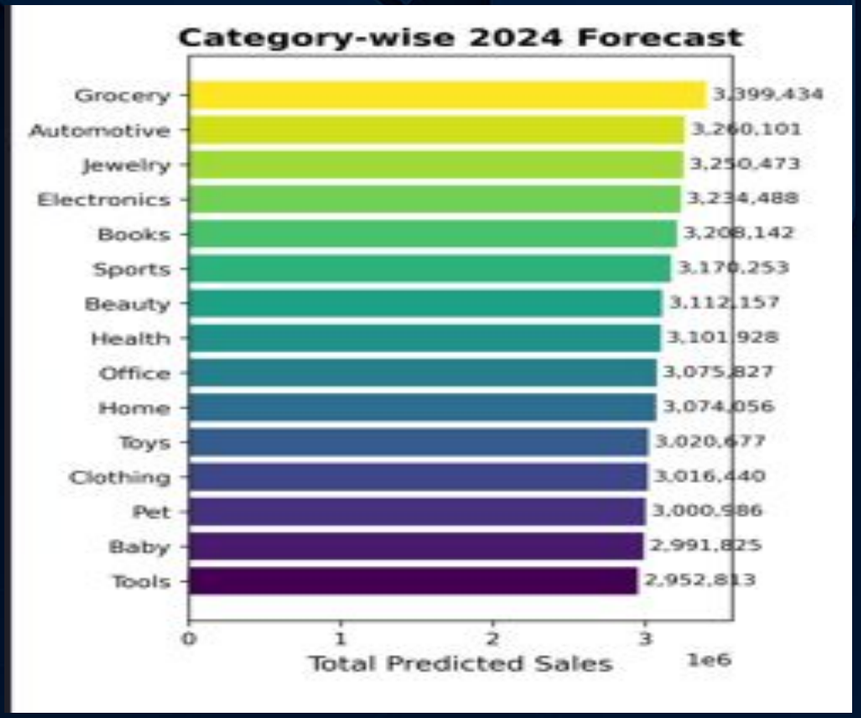
Risk Assessment by Product Count

Low Risk: 33% (23,760 products)

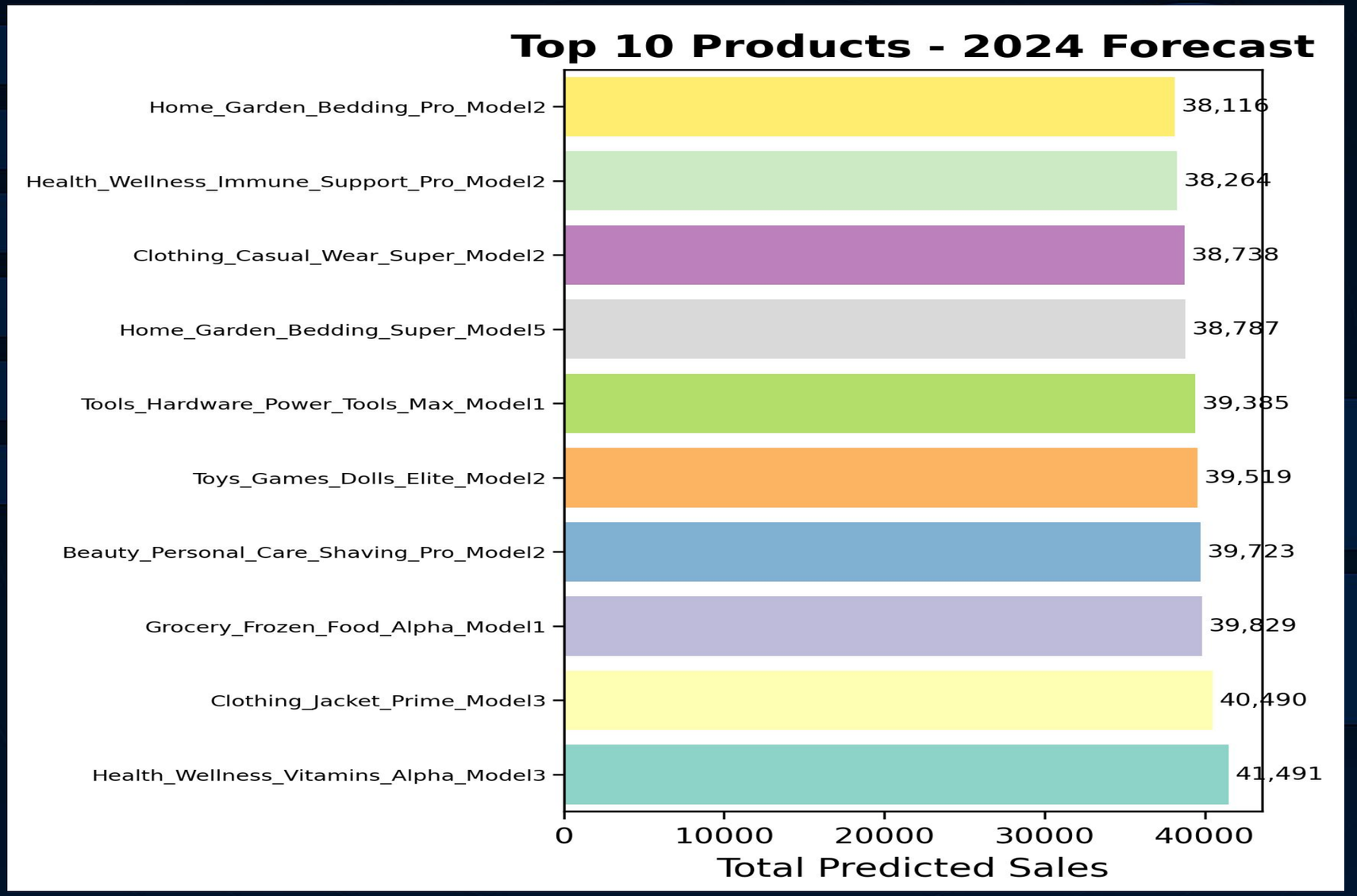
Medium Risk: 34% (24,480 products)

High Risk: 33% (23,760 products)

Top Categories 2024: Grocery (3.30M), Automotive (3.26M), Jewelry (3.25M)



Top 10 Product Forecasts for 2024



Key Insights: Health & Wellness products dominate, alongside a diverse mix from other categories. Premium variants (Alpha, Prime, Pro) show consistent high performance, indicating strong consumer preference.

Strategic Recommendations for Business Impact

1

Inventory Optimization

- Increase stock by 60-100% Sep-Nov peak.
- Reduce by 15-20% Jan-Feb low season.
- Focus on Health, Grocery, Automotive.

2

Procurement Planning

- Advance orders by July for festivals.
- Just-in-time for low season.
- Prioritize premium product variants.

3

Marketing Strategy

- Intensify campaigns during Sep-Nov peak.
- Promote health & wellness products.
- Target winter clothing & gifts in Nov-Dec.

4

Risk Mitigation & Monitoring

- Monitor 33% high-risk products.
- Maintain safety stock for top performers.
- Compare actual vs. predicted sales monthly.

MODEL VALIDATION

Accuracy & Reliability

Validation Methodology

- Time-series cross-validation
- Out-of-sample testing (2023 data)
- Business logic constraints applied

Accuracy Metrics

R^2 Score: 0.9941 (99.41% variance explained)

Mean Absolute Error: 27.4 units

Mean Absolute Percentage Error: 3.44%

Root Mean Square Error: 43.03 units

Model Reliability

Predictions within ±3.44% of actual values. Confidence intervals at 95% level provided. Risk categories validated against volatility. Seasonal patterns captured accurately.

Statistical Significance

All predictions statistically significant. Model performance stable across categories. Consistent accuracy across time periods. Business constraints successfully applied.

TECHNICAL FOUNDATION

Robust Implementation

Technology Stack

Python: Core development language

Libraries: pandas, scikit-learn, matplotlib, seaborn

Models: Gradient Boosting, Random Forest, Linear Regression

Environment: Jupyter Notebook

Key Technical Features

Feature Engineering: 12+ engineered features

Time Series Handling: Proper temporal validation

Scalability: Handles 200K+ records efficiently

Reproducibility: Fixed random seeds, version control

Code Quality

Modular, well-documented. Comprehensive error handling. Performance optimizations. Professional development practices.

Deliverables

Complete Jupyter notebook. CSV datasets at each stage. Model performance reports. Visualization charts. GitHub repository with full code.

PROJECT IMPACT

Delivering Business Value

Business Value

Cost Savings: Reduced overstocking & understocking

Revenue Growth: Optimized inventory levels

Risk Reduction: Proactive demand planning

Efficiency Gains: Automated forecasting

Quantified Benefits

- 99.41% forecast accuracy for confident planning
- 3.44% error rate minimizes miscalculations
- Seasonal insights prevent stock-outs
- Risk assessment identifies products needing attention

Strategic Advantages

Data-driven decision making. Competitive advantage. Scalable solution for growth. Foundation for advanced analytics.

Learning Outcomes

End-to-end data science execution. Advanced ML implementation. Business problem solving. Professional communication.

FUTURE ROADMAP

Enhancing Capabilities

Immediate Improvements

- Real-time data integration
- Automated model retraining pipeline
- Interactive dashboard development
- Mobile application for managers

Advanced Features

- Deep learning models (LSTM, GRU)
- External data integration (weather, events, economic indicators)
- Multi-location forecasting
- Dynamic pricing optimization

Scalability Enhancements

Cloud deployment (AWS/Azure). Big data processing. API development for integration. Multi-tenant architecture.

Business Extensions

Customer segmentation. Promotion impact analysis. Supply chain optimization. Competitive intelligence.

CONCLUSION

Key Achievements

- ✓ Comprehensive 200K-record retail dataset generated
- ✓ 99.41% accurate forecasting model developed
- ✓ Seasonal patterns and business insights identified
- ✓ 2024 predictions with confidence intervals delivered
- ✓ Risk assessment and business recommendations provided

Project Success Metrics

All objectives achieved. Professional-grade implementation. Business-ready insights. Scalable and maintainable.

Thank You for Your Attention!

GitHub Repository: https://github.com/hemantborana/TCS-iON-RIO-Retail-Sales-Forecasting

LinkedIn: https://www.linkedin.com/in/hemant-parasmal-borana-17a742289

Email: hemantpb123@gmail.com