ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

SALES INTELLIGENCE DASHBOARD WITH PREDICTIVE MODELING CAPABILITIES

Dr. A. Karunamurthy[1]*, R. Aravinth[2]
[1]Associate Professor, Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College (Autonomous), Madagadipet, Puducherry – 605107, India
[2]PG Student Department of MCA, Sri Manakula Vinayagar Engineering College (Autonomous), Madagadipet, Puducherry – 605107, India

ABSTRACT

This paper presents a Sales Intelligence Dashboard with Predictive Modeling Capabilities designed to convert raw transactional sales data into actionable business intelligence. The system combines descriptive analytics, predictive modeling, risk scoring, and interactive visualization in a single Streamlit-based application. The implemented analytical stack includes Linear Regression for baseline estimation, Random Forest for higher-capacity prediction, Prophet-based time-series forecasting for monthly trend projection, K-Means clustering for customer segmentation, Z-score analysis for anomaly detection, churn-risk scoring, lead scoring, and RFM analysis. In addition to analytics, the platform incorporates secure login, session continuity, dataset upload history, and module-wise dashboards for business users. The proposed solution is valuable because it does not stop with charts; instead, it interprets performance, identifies risk, and suggests where decision-makers should focus attention. Compared with isolated spreadsheet-based analysis, the dashboard offers a deeper, faster, and visually clearer workflow for sales monitoring and planning.

KEYWORDS: *Sales Intelligence, Predictive Modeling, Random Forest, Forecasting, Anomaly Detection, Churn Risk, Lead Scoring, RFM, Dashboard Analytics*

1. INTRODUCTION

Nowadays shops grab sale facts from bills, web purchases, updates in buyer accounts, discounted pricing, also actions logged at branch spots. While such data counts plenty, it often spreads across many workbook sheets - rarely used properly. People may know cash flow for the past month, but still struggle to see which section drives growth, which products might climb next, where shopper spending dipped recently, or which customers could stop showing up anytime soon.

This project sees sales figures in a new light. Beyond old transactions, they're hints pointing forward. A single platform gathers them - revealing performance, uncovering changes, estimating future outcomes, highlighting who matters right now. Quietly, the Sales Intelligence Dashboard with Predictive Modeling handles every piece. Ahead is where numbers start making sense, not behind. The real work kicks in once the usual report wraps up.

A single aim drove the design - breaking the tool into pieces that make data easier to understand. Instead of crowding everything onto one busy display, it uses separated sections: Main Dashboard, Analytics Module, Predictions Module, Churn Risk Analysis, Lead Scoring, RFM Analysis, Reports, Settings, along with an AI Chatbot. Every section does just one job, much like different magnifiers for unique problems. People leading sales teams often pay closest attention to targets and how numbers stack up across areas. Odd time periods reveal how models behave differently. Because patterns change, analysts pay close attention. Unstable accounts catch marketing eyes, also promising leads. When things are split out instead of lumped, finding your way takes less effort. Layout gives clues, yet purpose shapes placement. What each person uses affects where it sits.

What you mean needs to show up clearly. Many learners busy themselves shaping algorithms while skipping why they work, or crafting visuals that shine but ignore actual use. This place? It slows down on purpose to walk through every piece out loud. Start with the goal sitting up front - next, where numbers come from, how steps unfold, what outcomes aim to become, lastly how choices may tilt after seeing them. This setup changes the file from mere jots into a version that invites feedback. Because connections form between every part, clarity arrives without effort.

2. PROBLEM STATEMENT AND RESEARCH MOTIVATION

Figures crowd every sales report - monthly swings, top items, overall counts dominate. But teams require more than frozen moments on a page. Trouble might hide behind a sudden dip, one that strikes early. Value shows up when income gets forecast ahead of quarter's end. Certain customers matter more; recognizing them reshapes choices. Fewer cracks show when you inspect solid parts closely, unlike weaker areas. Moves made too soon on uncertain ground change what happens later.

A lone system gets pushed to its edge. When messy sales numbers flood in, can it juggle several ways of learning at once? Meaning must snap into place instantly. Not just charts. Real clarity for

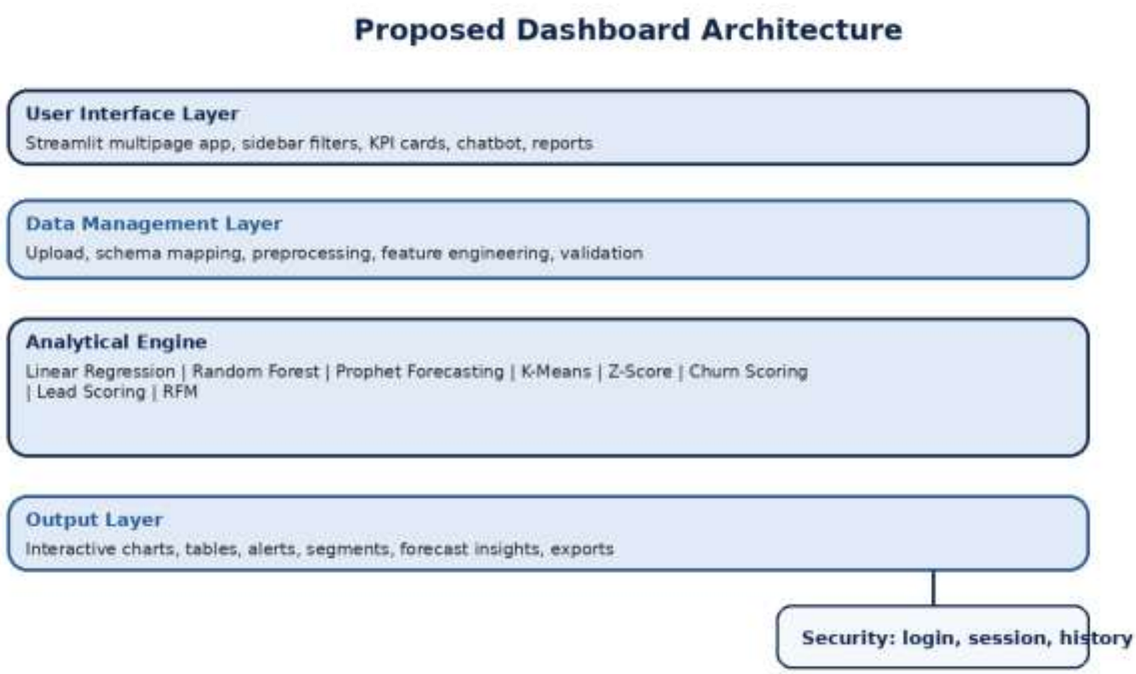
ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

people who do not touch programming. The pressure is on. Pictures speak before words get a chance. Though deep beneath, numbers bend in quiet patterns, wisdom moves without teaching.

Most modern tools skip a key piece of what drives this work. Instead of only number-heavy predictions, there's now insight into which clients may slip away. While sorting users by actions is common, guessing their next dollar remains tough. Even when dashboards show clean outputs, strange trends or deeper sequences go unnoticed. A single platform has started merging these scattered abilities at last.

3. OBJECTIVES OF THE PROPOSED SYSTEM

Objective	Technical Approach	Business Value
Generate clear sales summaries	KPI cards, charts, category and regional aggregation	Quick executive overview
Forecast upcoming revenue	Linear Regression, Random Forest, Prophet	Planning and target setting
Detect suspicious behavior	Z-score anomaly detection	Early warning for spikes or drops
Segment customers intelligently	K-Means and RFM analysis	Better retention and marketing
Measure relationship risk	Churn-risk scoring	Customer recovery actions
Prioritize opportunities	Lead-scoring framework	Focus on likely conversions
Support continuous usage	Authentication and upload history	Secure personalized access



ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Figure 1. Proposed dashboard architecture showing layered system design.

Figure 2. Operational workflow from login and upload to analytics and reporting.

4. DATASET DESIGN AND DATA PREPARATION

A tidy start helps you find numbers faster. When things get messy? This setup keeps control - dates show first, products appear next, categories arrive after that, regions follow close behind, sales wrap up the core set. After those, orders step in, customer age pops next, returning visits trickle later, conversion stats sneak beneath, closing with one unique buyer ID at the end. Words like revenue act exactly as sales do - their role does not shift. Some cells may sit empty sometimes, headings can seem odd, but alignment holds regardless. Below, words such as income or profit slowly collect behind a single title. Even when names shift, the meaning remains unchanged. Where one part ends, another begins - no force required. The way it sounds does not alter its sense. Quiet keeps pieces linked, despite different forms.

One second at a time, scattered bits throw off forecasts. Right after alignment kicks in, timestamps snap into place. Numbers shift shape next - morphed for machine ease. When gaps show up, clever estimates patch them, never leaving blanks behind. A single number lifted from a clock can unfold slowly, stretching through weeks, sometimes drifting into different times of year -

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

hiding just beyond seconds we thought were gone. Shift the angle slightly, suddenly shapes appear that existed all along but stayed out of sight. The way we line up hours opens doors to things that sat still, unnoticed, beneath the surface. When digits change far enough, circular years begin showing themselves, stepping into view.
A shape begins before sight catches up. From remnants left over, movement emerges piece by piece. It keeps moving, even when empty space remains. When seen clearly, fragments link without help. A shape changes its width by itself, adjusting uneven parts. While one flow settles, another follows close behind, syncing up naturally. Already back again? Numbers appear fresh, shifting right after they arrive, holding totals firm. From a choppy beginning - rough inputs weave softly into smooth sections since the structure bends but does not break.

4.1 Core Input Fields

Field	Type	Usage in Models	Interpretation
Date	Date	Forecasting, Anomalies, Seasonality	When Transactions Happened
Sales	Numeric	Regression, Forecasting, Anomaly Detection	Revenue Signal
Orders	Numeric	Lead Scoring, Regression, Churn Scoring	Demand Intensity
Region	Categorical	Regional Charts, Model Encoding	Geographic Contribution
Product / Category	Categorical	Performance Ranking, Model Encoding	Product Mix
Repeat Customer	Binary	Churn, Lead Scoring, Segmentation	Retention Behavior
Conversion Rate	Numeric	Lead And Churn Scoring	Sales Effectiveness
Customer Age / Profile	Numeric	Lead Scoring, Segmentation	Customer Characterization

5. SYSTEM ARCHITECTURE AND MODULE OVERVIEW

One piece sits on top of another, like steps going down. At the front, Streamlit shows pages people click through, with side menus, key stats up front, plus a bot that answers questions. Data comes in, then gets checked right away - structure confirmed before anything else happens. Features get built during cleanup, shaping raw inputs into something usable. Inside, models run their course while scores for decisions are calculated quietly. What emerges lands in visuals, warnings pop up when needed, grids appear, files wait to be saved. Last part delivers it all, clear and ready.

A single screen packed with every chart might seem efficient, yet it often leads to clutter that hampers understanding. Spreading features over separate views helps keep things clear. When each model lives on its own page, navigation follows a smoother path from big picture to specifics. This setup works well since the dashboard tracks multiple systems at once.

5.1 Major Modules

- Main Dashboard – Displays headline KPIs such as total sales, top product, best region, and overall trend direction.
- Analytics Module – Examines monthly patterns, category contribution, product performance, and regional distribution.
- Predictions Module – Compares actual sales with projected sales and visualizes future trends with confidence bands.
- Anomaly Detection Module – Highlights unusual months using Z-score based thresholds so that abnormal changes are not ignored.
- Lead Scoring Module – Assigns a commercial priority score based on conversion quality, order intensity, and customer profile.
- Churn Risk Module – Estimates which customers or segments require retention efforts.
- RFM Analysis Module – Segments users into Champions, Loyal Customers, Potential Loyalists, At Risk, and Lost groups.
- Reports and Settings – Support export options, threshold tuning, and user preferences.
- AI Chatbot – Allows business users to ask natural-language questions about the uploaded sales data.

6. DEEP EXPLANATION OF THE MACHINE LEARNING AND SCORING MODELS

6.1 Linear Regression: Baseline Forecasting Model

Linear Regression is the simplest predictive model in the system and acts as a baseline. It assumes that sales can be explained as a weighted combination of input variables. In this project, these inputs may include order count, conversion rate, repeat-customer behavior, month number, quarter, and encoded categorical information such as region or product category.

The importance of this model is not only its prediction. Linear Regression is easy to interpret because each coefficient shows the direction and approximate strength of association between an input feature and the sales output. If the coefficient of orders is positive, increasing orders tends to increase predicted sales. If a discount variable is included and its coefficient is negative, higher discounts may be associated with lower realized revenue.

Although business data is often nonlinear, a baseline model remains useful because it provides a transparent benchmark. When compared against stronger models such as Random Forest, it helps prove whether added complexity genuinely improves performance.

6.2 Random Forest: Higher-Capacity Sales Prediction

Random Forest is an ensemble model built from multiple decision trees. Each tree learns a different view of the dataset by sampling observations and features. The final output is the average prediction across all trees. This averaging process makes the model less sensitive to noise and usually more accurate than a single decision tree.

In the context of this dashboard, Random Forest is especially effective because sales are influenced by interactions that are not perfectly linear. For example, the impact of region may depend on the product category, and the effect of conversion rate may change across customer segments. Ensemble trees can capture these interactions without requiring the developer to manually specify them.

Another important advantage is feature importance. Random Forest can estimate which variables contribute most strongly to predictive accuracy. In a business explanation, this helps answer practical questions such as whether time, region, product, or repeat purchase behavior is most informative for future sales.

6.3 Prophet-Based Time Series Forecasting

Sales forecasting over time is different from general regression because the order of observations matters. Prophet, a decomposition-based time-series model, is useful when monthly revenue exhibits trend and seasonality. It separates the series into interpretable parts such as long-term direction, repeating patterns, and irregular noise.

In the dashboard, monthly sales are aggregated from raw transactions before forecasting. This is important because decision-makers usually monitor weekly or monthly totals rather than individual invoice-level rows. Once aggregated, the model can project future months and display both the central estimate and the uncertainty band.

From a business perspective, this module supports budgeting, staffing, stock planning, and campaign scheduling. A rising forecast may justify additional inventory or promotional investment. A declining forecast may prompt discount control, territory review, or lead generation campaigns.

6.4 K-Means Clustering: Customer Segmentation

K-Means is an unsupervised learning method used when the system must group customers without predefined labels. The algorithm repeatedly assigns each customer to the nearest centroid and updates centroid positions until the clusters stabilize. In this project, clustering can use features such as sales value, purchase frequency, conversion rate, and repeat behavior.

The major benefit of K-Means is that it converts a large undifferentiated customer base into interpretable segments. One cluster may contain high-value repeat customers, another may capture price-sensitive low-frequency buyers, and a third may represent developing customers with moderate conversion potential.

After clustering, the dashboard uses scatter plots and summary labels to help users understand segment structure. This makes segmentation actionable rather than purely mathematical.

6.5 Z-Score Anomaly Detection

Anomaly detection is essential because averages often hide abnormal events. The dashboard uses Z-score analysis to measure how far each monthly sales value deviates from the mean in standard-deviation units. When a score crosses a predefined threshold, the system flags that month as unusual.

This method is especially effective in business situations where sudden spikes or drops may signal campaign success, reporting errors, stock issues, demand shocks, or process breakdowns. A single abnormal month can distort managerial interpretation if it is not highlighted explicitly.

The anomaly charts in the project do more than compute a number. They visually place sales relative to a band, making it easy to see when normal variation ends and abnormal behavior begins.

6.6 Churn Risk Scoring

Churn risk is not directly observed in many sales datasets, especially when there is no formal subscription cancellation label. Therefore, the project uses a composite score. Recency contributes positively to risk when the customer has not purchased recently. Frequency, monetary value, repeat behavior, and conversion efficiency are used inversely because strong engagement implies lower risk.

The resulting score is mapped into categories such as High Risk, Medium Risk, and Low Risk. This translation from continuous number to managerial category is important because business teams act more easily on simple groups than on raw decimals. High-risk accounts can be targeted with follow-up calls, offers, or service checks.

This module demonstrates how domain knowledge can be combined with analytics. Instead of waiting for churn to happen, the system estimates relationship weakness before revenue is lost.

6.7 Lead Scoring

Lead scoring ranks opportunities by expected commercial value. In the project, the score is constructed from factors such as conversion rate, repeat tendency, normalized order volume, and customer-profile fit. The purpose is to help a sales team allocate limited attention more effectively.

A good lead-scoring module does not merely label everyone as medium priority. It must produce separation: the best opportunities should visibly stand out. The dashboard therefore visualizes distribution by priority and provides a detailed scoring table where products or segments can be compared side by side.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

This is strategically useful because sales performance depends not just on how many leads exist, but on whether the right leads are pursued at the right time.

6.8 RFM Analysis

RFM stands for Recency, Frequency, and Monetary value. It is a classical but still highly effective framework for customer value segmentation. Recency measures how recently a customer purchased. Frequency measures how often purchases occur. Monetary measures how much value the customer generates.

In this dashboard, each dimension is converted into a score and summed to form a total RFM score. The final labels such as Champions, Loyal Customers, Potential Loyalists, At Risk, and Lost make the result intuitively understandable. This helps bridge the gap between analytics and action. For example, Champions should be retained and rewarded, At Risk users should be reactivated, and Lost users may require special campaigns or may be deprioritized depending on cost.

RFM is especially valuable because it gives a business-friendly interpretation without requiring heavy black-box modeling. It complements machine learning by offering an explainable segmentation layer.

7. Dashboard Screens and Result Interpretation

The practical value of the project is best understood through the dashboard pages themselves. The following figures present selected screens from the implemented application. Each screen corresponds to a distinct business question and demonstrates how the system converts model output into readable decision support.

The screenshots included below are not decorative additions. They are evidence of module integration: the forecast charts, anomaly panels, customer segments, risk distributions, and AI query interface all show that the project works as a usable analytical platform rather than as a collection of disconnected algorithms.

Figure 3. Monthly sales trend and region-wise sales distribution on the main dashboard.

This figure summarizes the main business overview. The line chart helps the viewer understand how sales move month by month, while the donut chart shows regional contribution. Together, these visuals answer two foundational questions: how is revenue changing over time, and where is that revenue coming from?

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

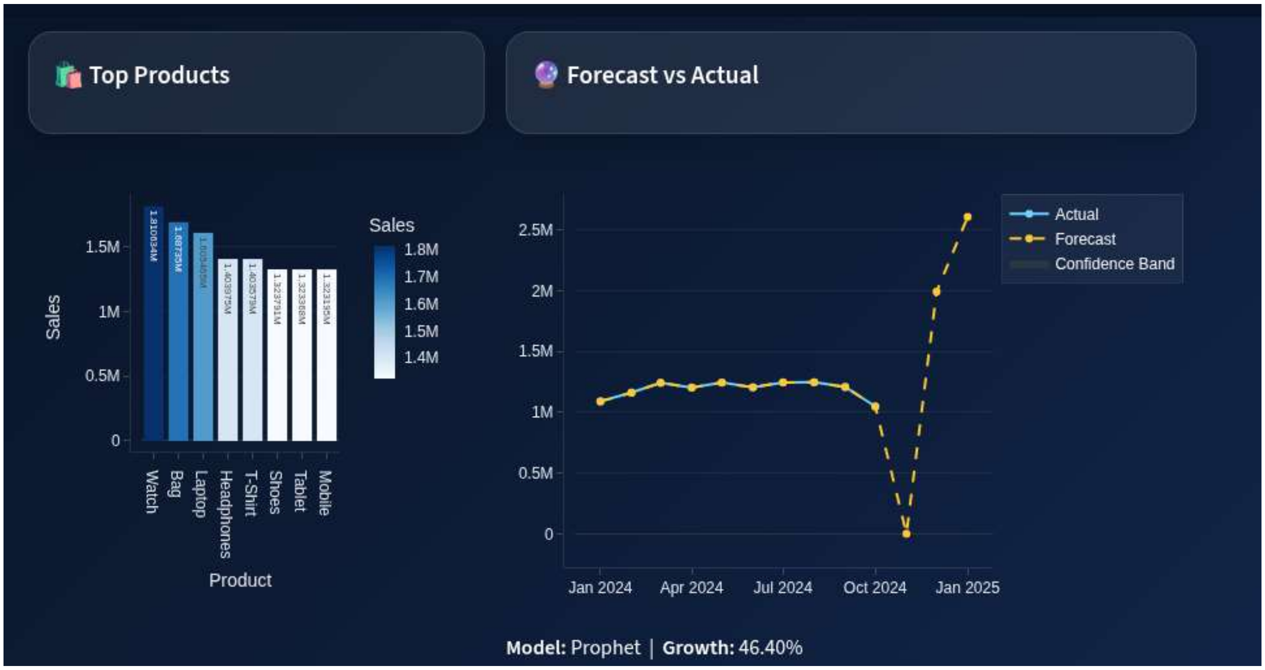


Figure 4. Top products and forecast-versus-actual visualization.

The left-side bar chart ranks products by revenue contribution, while the right-side plot compares realized values against the forecast path. This combination is useful because management often needs both current performance leadership and near-future direction in the same analytical context.

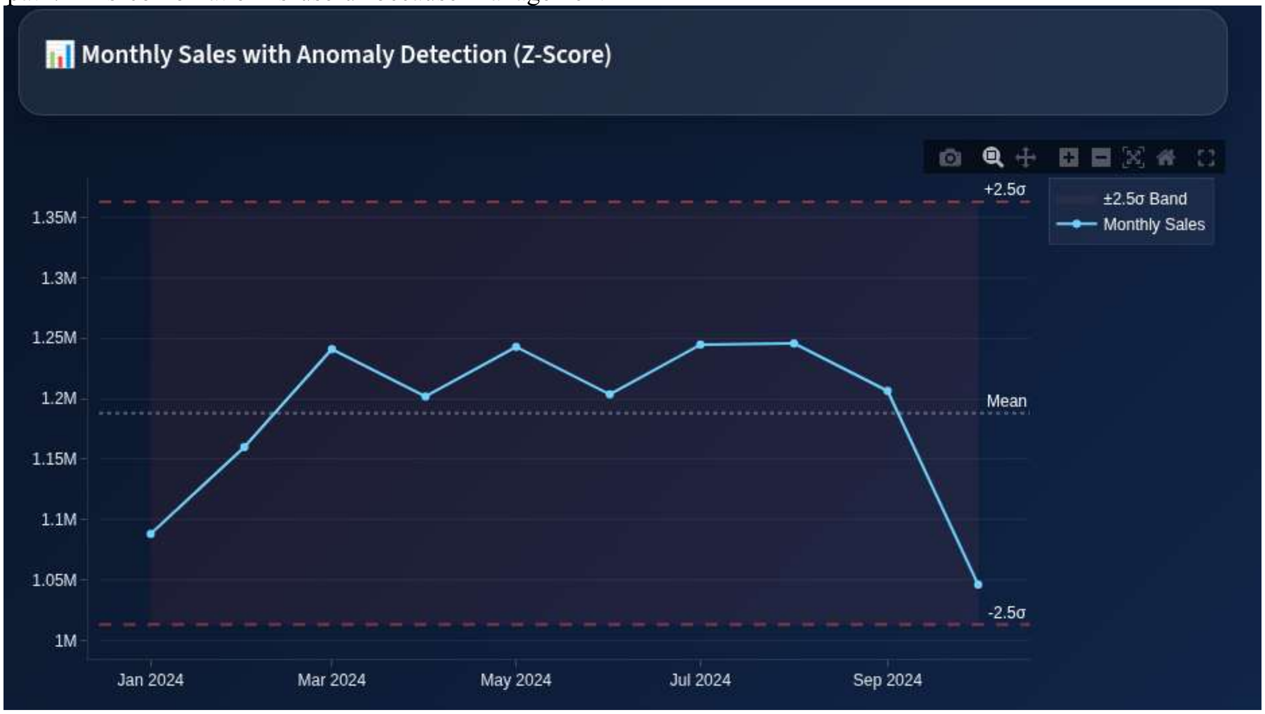


Figure 5. Monthly sales anomaly detection using a Z-score band.

The anomaly chart provides interpretive depth beyond a normal sales line. The dotted mean and threshold band make it immediately visible when a monthly value behaves abnormally. This allows users to investigate operational or market reasons behind unusual movement.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Figure 6. Churn risk distribution and churn score histogram.

This page translates customer weakness into business categories. Instead of only showing raw customer data, it groups behavior into risk tiers and reveals how the score is distributed. Such a view helps prioritize retention actions and understand whether churn risk is concentrated or widespread.

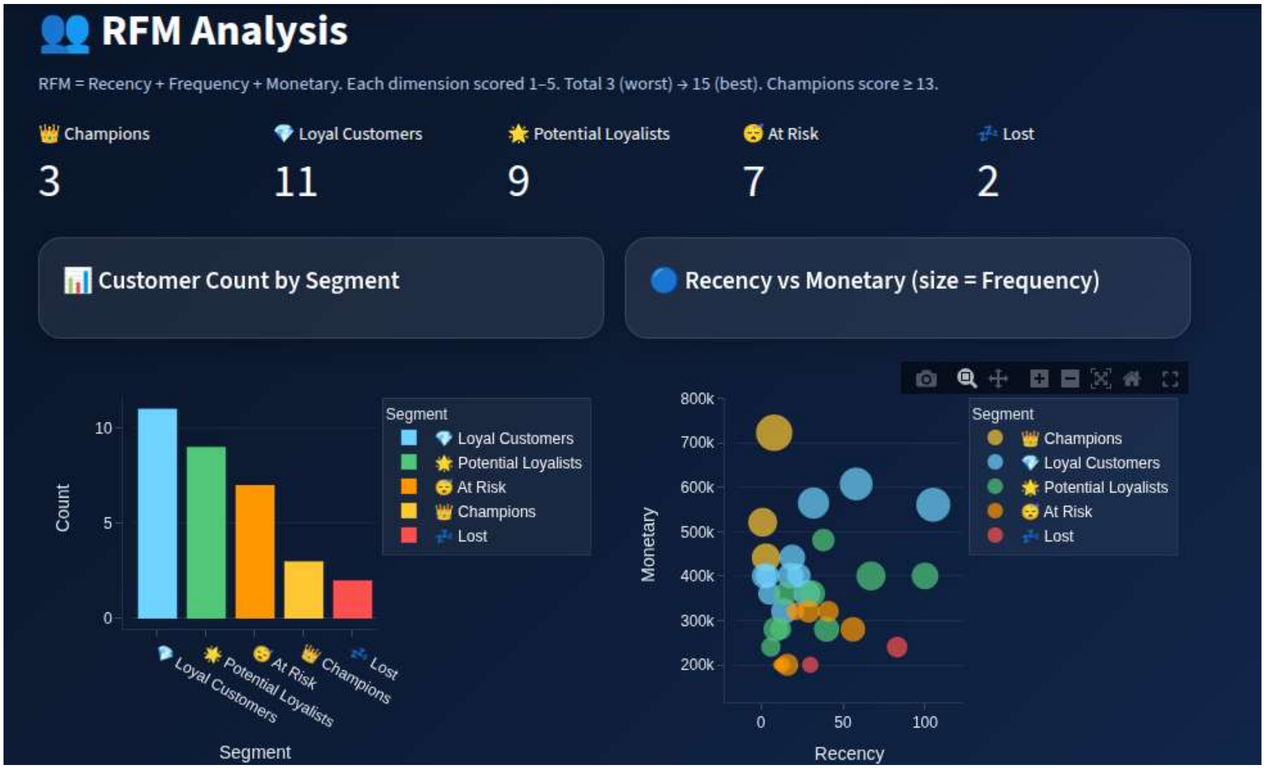


Figure 7. RFM segment counts and recency–monetary–frequency visualization.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

The RFM page explains customer quality through a business lens. Count bars show how many customers fall into each segment, and the bubble plot provides a richer relationship between recency, monetary value, and frequency. This is extremely useful for segmentation strategy.

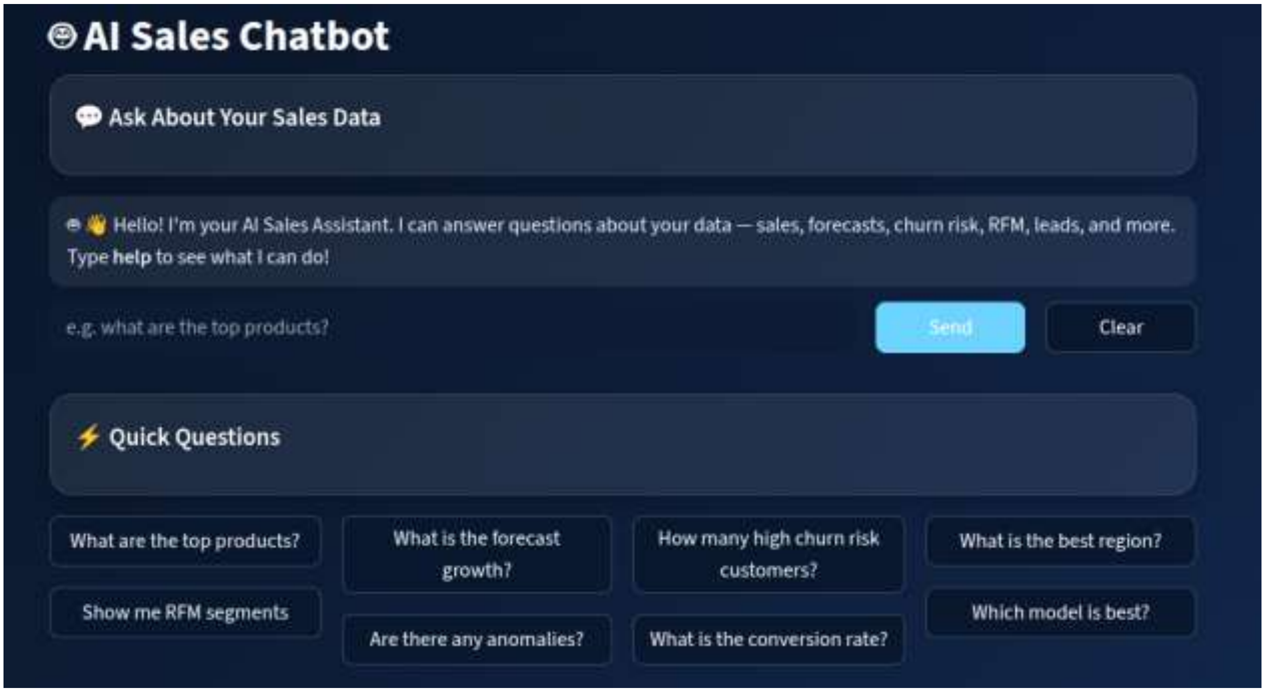


Figure 8. AI sales chatbot interface for natural-language interaction.

The chatbot page expands dashboard usability. Users who do not want to manually navigate every chart can ask direct questions such as top products, forecast growth, best region, conversion rate, or churn-risk counts. This feature improves accessibility for non-technical stakeholders.

8. MODEL COMPARISON AND BUSINESS INTERPRETATION

A dashboard with several models must clearly explain why each model exists. The table below summarizes the role of each analytical component, the type of output it produces, and the business interpretation users should derive from it.

The point is not to claim that one model replaces all others. In fact, the strength of the project lies in complementarity. Regression predicts quantity; clustering reveals structure; anomaly detection identifies exceptions; scoring modules prioritize action; and RFM translates customer history into a managerial segmentation framework.

Model / Method	Primary Input	Output	Managerial Meaning
Linear Regression	Time and behavioral features	Baseline predicted sales	Transparent benchmark and directional relationships
Random Forest	Mixed numeric + categorical inputs	Improved predicted sales	More accurate nonlinear forecasting
Prophet	Monthly aggregated sales	Future time-series path	Planning, budgeting, seasonality tracking
K-Means	Customer behavior features	Cluster label	Segment-specific marketing strategy
Z-Score	Monthly sales totals	Anomaly flag	Investigate exceptional months
Churn Scoring	Recency, frequency, value, repeat, conversion	Risk category	Retention priority
Lead Scoring	Conversion, orders, fit, repeat behavior	Priority score	Sales follow-up ranking
RFM	Recency, frequency, monetary value	Value segment	Customer lifecycle strategy

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

9. IMPLEMENTATION STACK

The project is implemented primarily in Python. Pandas is used for ingestion, cleaning, aggregation, and feature engineering. Scikit-learn supports the machine learning models such as Linear Regression, Random Forest, scaling, PCA, and K-Means clustering. Prophet is used for time-series forecasting. Plotly and Streamlit together provide interactive visuals and page navigation.

From a software-engineering perspective, the project benefits from modular structure. Authentication logic, data loading, dashboard pages, prediction modules, scoring modules, and report generation can be maintained independently. This improves extensibility. New models such as XGBoost or LSTM could be added later without redesigning the entire interface.

10. STRENGTHS OF THE PROPOSED PROJECT

- Integrates multiple analytical models in one business-oriented application.
- Explains sales from descriptive, predictive, and risk-based viewpoints.
- Improves user clarity through separate modules instead of one overloaded page.
- Supports action through churn categories, lead priorities, and RFM segments.
- Includes actual interface implementation with charts, tables, and AI query support.
- Is flexible enough for academic demonstration and practical prototyping.

11. LIMITATIONS AND SCOPE FOR IMPROVEMENT

As with most analytics systems, performance depends on data quality. If columns are missing, mislabeled, or highly noisy, model quality can decline. Forecasting quality also depends on time depth; a very short history limits the reliability of seasonality estimation.

The current project is highly suitable as a strong student or prototype business-intelligence system, but an enterprise deployment would require additional production-level controls such as database persistence, role-based authorization, model monitoring, and API integration with CRM or ERP systems.

12. CONCLUSION

The Sales Intelligence Dashboard with Predictive Modeling Capabilities demonstrates how a single application can bring together performance monitoring, predictive analytics, customer segmentation, risk detection, and business prioritization. The value of the project lies in integration. Instead of giving users only static charts or only model outputs, it creates a complete decision-support environment.

For project evaluation, this work is strong because it contains both technical depth and business interpretation. Each model is not only implemented but also placed in a meaningful managerial context. The dashboard therefore functions as an intelligent analytical assistant rather than as a simple reporting screen.

In summary, the proposed system transforms ordinary sales records into a structured narrative about business performance: what happened, why it may have happened, what is likely to happen next, and where the organization should focus attention.

13. FUTURE ENHANCEMENT DIRECTIONS

- Use deeper forecasting models such as LSTM or Temporal Fusion Transformer for richer sequence learning.
- Introduce real-time streaming so dashboards update as new transactions arrive.
- Connect the system to a database back end such as PostgreSQL for persistent multi-user storage.
- Add explainable AI components such as SHAP values for model interpretability.
- Expand the chatbot to support narrative summaries and automated report generation.
- Deploy the platform to cloud infrastructure for broader access and scalability.

14. SUGGESTED REFERENCES

1. *L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.*
2. *T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer, 2009.*
3. *F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.*
4. *W. McKinney, Python for Data Analysis. O'Reilly Media, 2022.*
5. *S. J. Taylor and B. Letham, "Forecasting at Scale," The American Statistician, vol. 72, no. 1, pp. 37–45, 2018.*
6. *Streamlit Documentation, "Build data apps in pure Python."*
7. *P. Kotler and K. L. Keller, Marketing Management. Pearson.*
8. *Business intelligence and customer analytics references may be expanded according to the target journal style.*