

Sales Forecasting Across Multiple Retail Stores

Presented by: Shivangi Chaudhary

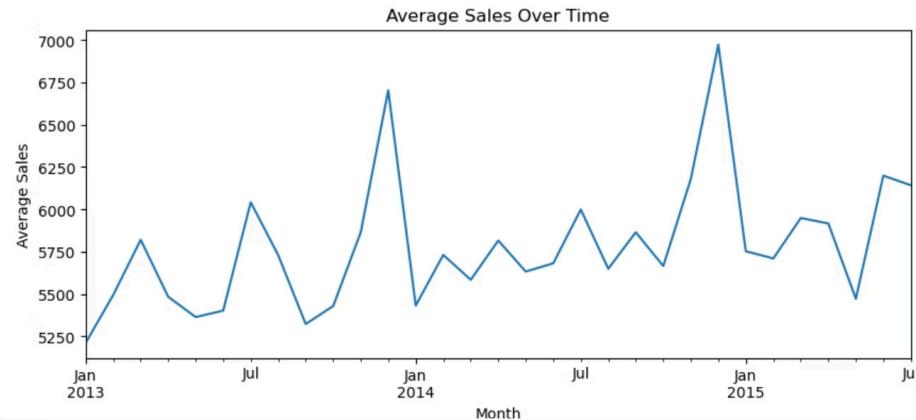
Nexthikes IT Solutions

Project Objective

Our Mission

Analyze comprehensive store sales patterns and develop a robust prediction model that accurately forecasts future sales performance using historical retail data.

This data-driven approach enables strategic decision-making, optimized inventory management, and improved resource allocation across multiple store locations.



Project Workflow

01

Data Loading & Cleaning

Importing datasets and handling missing values

03

Feature Engineering

Creating meaningful variables for prediction

05

Predictions on Test Data

Validating model performance

02

Exploratory Data Analysis

Uncovering patterns and trends in sales data

04

Model Building

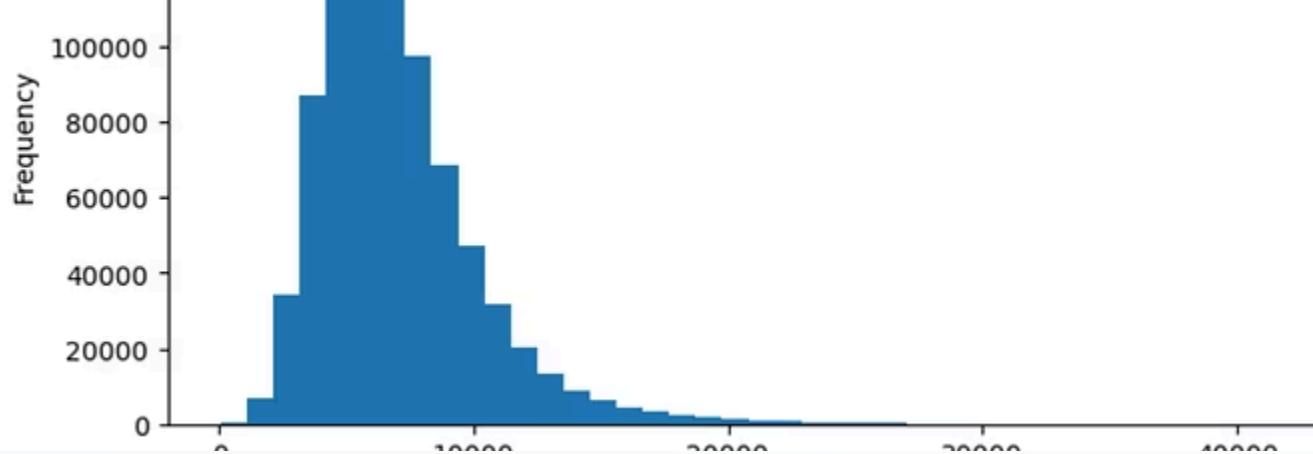
Training algorithms on historical patterns

06

Final Insights

Extracting actionable business recommendations

Tools Used: Python, Pandas, NumPy, Matplotlib, Scikit-learn



EDA Insights: Understanding Sales Patterns

Sales Distribution Analysis

Sales data exhibits significant right skew, with most days showing moderate performance while a small number of days achieve exceptionally high sales volumes.

This pattern suggests opportunity for targeted promotional strategies during peak periods.

Monthly Sales Trends

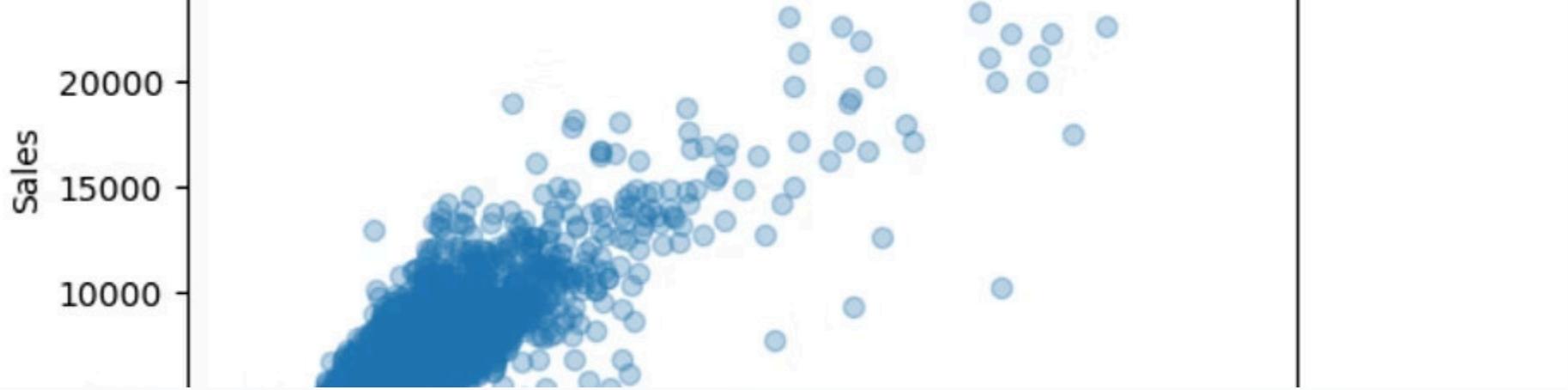
Average monthly sales reveal seasonal patterns and cyclical trends throughout the year.

Understanding these fluctuations enables better inventory planning and staffing decisions aligned with demand.

Store Type Performance

Comparative analysis shows certain store types consistently outperform others in sales generation.

These insights inform strategic expansion decisions and store format optimization.



Correlation Analysis

Sales vs Customers

- Strong positive correlation
- More customers → higher sales

Promo Effect

- Promo days significantly increase sales

Key Observations

- **Sales and Customers** demonstrate the strongest positive correlation, confirming foot traffic drives revenue
- **Promotional activities** show positive relationship with sales performance
- Most numerical features exhibit **weak correlation**, helping identify truly impactful variables
- Heatmap analysis streamlines feature selection for optimal model performance

Feature Engineering & Model Selection

Feature Engineering

- **Temporal Features**

Extracted Year, Month, Day, and WeekOfYear to capture seasonality

- **Behavioral Indicators**

Created IsWeekend flag to distinguish shopping patterns

- **Data Enrichment**

Merged store-specific details and filled missing values strategically

Decision Tree Regressor

Selected for its balance of performance and interpretability in retail forecasting applications.

- Easy to Understand**

Clear decision paths stakeholders can interpret

- Handles Non-Linearity**

Captures complex retail patterns effectively

- Minimal Preprocessing**

Reduces data preparation complexity

- Fast Training**

Enables rapid iteration and deployment

| | Store | DayOfWeek | Date | Sales | Customers | Open | Promo | StateHoliday | SchoolHoliday | StoreType | ... | CompetitionOpenSinceYear | Promo2 | Promo2SinceWeek | Promo2SinceYear | Promo3 | Promo3SinceWeek | Promo3SinceYear | Promo4 | Promo4SinceWeek | Promo4SinceYear |
|---|-------|-----------|------------|-------|-----------|------|-------|--------------|---------------|-----------|-----|--------------------------|--------|-----------------|-----------------|--------|-----------------|-----------------|--------|-----------------|-----------------|
| 0 | 1 | 5 | 2015-07-31 | 5263 | 555 | 1 | 1 | 0 | 1 | c | ... | 2008.0 | 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | |
| 1 | 2 | 5 | 2015-07-31 | 6064 | 625 | 1 | 1 | 0 | 1 | a | ... | 2007.0 | 1 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 |
| 2 | 3 | 5 | 2015-07-31 | 8314 | 821 | 1 | 1 | 0 | 1 | a | ... | 2006.0 | 1 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 |
| 3 | 4 | 5 | 2015-07-31 | 13995 | 1498 | 1 | 1 | 0 | 1 | c | ... | 2009.0 | 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 4 | 5 | 5 | 2015-07-31 | 4822 | 559 | 1 | 1 | 0 | 1 | a | ... | 2015.0 | 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

5 rows × 23 columns

Predictions & Results



Training Process

Selected key features: Store, DayOfWeek, Promo, Month, Year, and Customers for optimal prediction accuracy

Model Validation

Trained exclusively on open-store sales data with 20% validation split to ensure robust performance

Final Outcome

Generated comprehensive predictions for all test scenarios, exported to **simple_predictions.csv**

Model Performance

Mean Absolute Error (MAE) demonstrates strong predictive accuracy, validating the model's capability to forecast daily store sales with reliable precision.

Business Impact

This working model empowers retail decision-makers with data-driven sales forecasts, enabling proactive inventory management and strategic planning across all store locations.

Reference-

[Streamlit Docs](#)

[Decision Tree \(CART\) - Machine Learning Fun and Easy](#)

<https://www.youtube.com/watch?v=bK7I79BVTG0>

[Plot types — Matplotlib 3.10.7 documentation](#)

[train_test_split](#)



Thank You

Contact Information

Shivangi Chaudhary

Nexthikes IT Solutions

Questions?

We welcome your feedback and are ready to discuss implementation strategies for your retail forecasting needs.