

**The fundamentals of artificial intelligence  
  
Subject: Sales forecasting**

**Vuqar Aliyev**

**Elgun Mammadsatov**

**Ismat Samadov**

**Ramil Valiyev**

**Dey Derabati Roy  
2024/2025 Fall semester  
First course  
Master business administration  
22\_24\_03\_E27-24  
  
2024**

# Plan

1. Introduction  
2. Dataset Overview  
3. Methodology  
4. Model and Metrics  
5. Results  
6. Conclusion  
7. Literature

**Introduction**

Sales forecasting is critical for businesses. Accurate forecasts help them make data-driven decisions on resource allocation, production, and budgeting. It can optimize supply chains, manage inventory, predict demand, and set sales targets.

Without reliable forecasting, companies may face stockouts, overproduction, or underutilization of resources, which can hurt profitability. Leveraging historical data and predictive models can mitigate these risks and align operational strategies with anticipated market trends.

In particular, sales forecasting provides key benefits such as:

* Inventory Optimization
* Demand Planning
* Resource Allocation
* Revenue Targets and Financial Planning

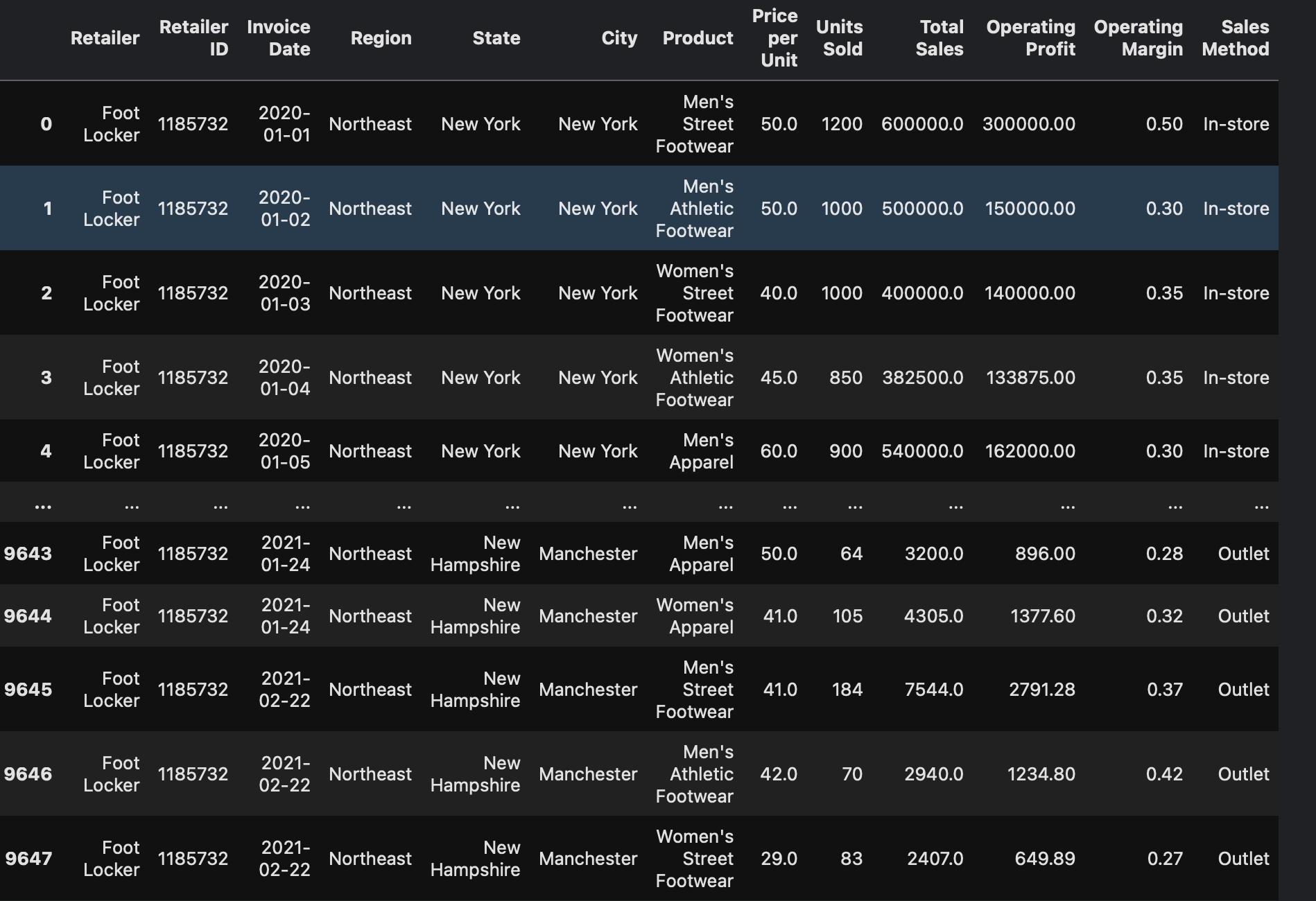
This project focuses on Adidas US sales data to predict total sales using features like Price per Unit, Units Sold, and Operating Profit. By analyzing these features, we aim to uncover sales patterns and trends.

We will create a Linear Regression model to predict future sales based on historical trends. Linear regression is well-suited for this task as it provides simplicity and interpretability.

By delivering accurate sales forecasts, the project enables Adidas US to align its strategic planning with data-driven insights, ensuring more efficient operations, optimized inventory levels, and improved financial outcomes.

# Dataset Overview

The Adidas US Sales Dataset contains data related to sales transactions, including Price per Unit, Units Sold, Total Sales, and categorical variables such as Retailer, Region, and Product. Several preprocessing steps were performed: converting Invoice Date into time-based features, one-hot encoding categorical variables, and standardizing numerical features using StandardScaler.



# Methodology

## Data preprocessing involves transforming raw data into a clean, structured, and useful format for modeling. Key steps include data cleaning, data transformation, data reduction, and handling imbalanced data. Without proper preprocessing, models may be inaccurate, biased, or unable to converge. Linear regression, chosen for sales forecasting, models the relationship between one or more independent variables and a continuous dependent variable. It's simple, interpretable, efficient, and serves as a baseline model for comparison. Data splitting into training and testing sets ensures model evaluation on unseen data, preventing overfitting.

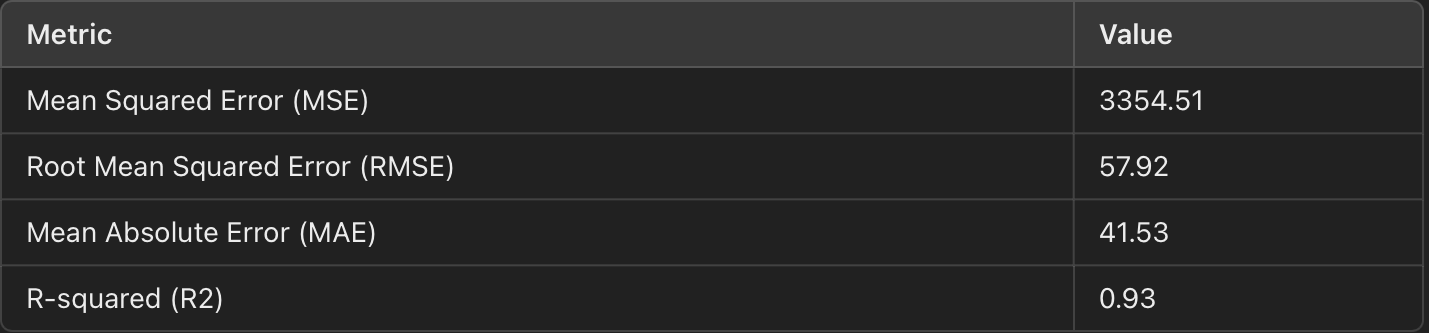
## 

## Model and Metrics

**Linear Regression:**

Linear regression is a statistical method that is used to determine the relationship between one or more independent variables and a dependent variable. It is a simple and effective way to model and predict the outcome of a particular event. Linear regression is based on the assumption that the relationship between the independent and dependent variables is linear, meaning that the dependent variable changes in a constant and predictable way as the independent variable changes.

**Main 4 Metrics:**

1. **Mean Squared Error (MSE):** MSE is a measure of how far, on average, the predicted values are from the actual values. It is calculated by summing the squared differences between the predicted and actual values, and then dividing by the number of observations.
2. **Root Mean Squared Error (RMSE):** RMSE is the square root of MSE. It is a measure of the average magnitude of the errors.
3. **Mean Absolute Error (MAE):** MAE is the average of the absolute differences between the predicted and actual values. It is a measure of the average size of the errors.
4. **R-squared:** R-squared is a measure of how well the model fits the data. It is calculated by dividing the explained variance by the total variance. The explained variance is the variance of the predicted values, and the total variance is the variance of the actual values.

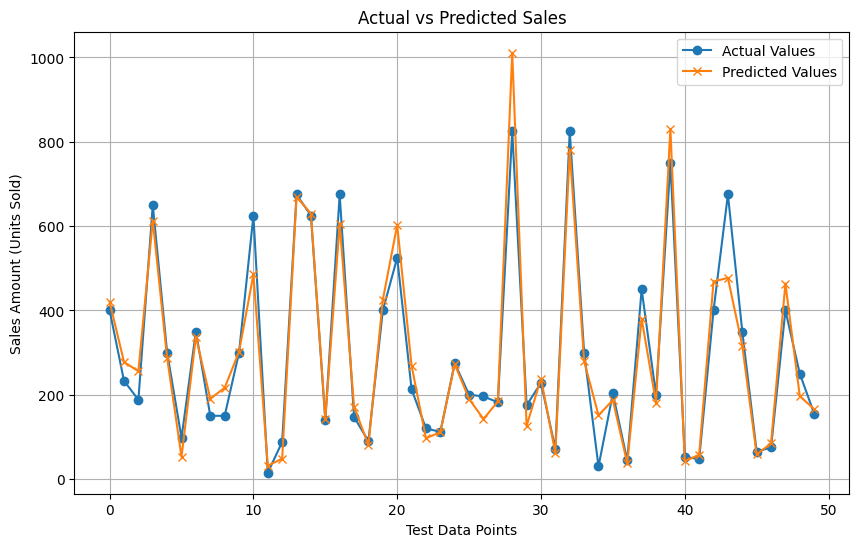
# Results

The performance of the **Linear Regression model** was assessed using four key metrics to ensure a comprehensive evaluation of its predictive capabilities. These metrics provide insights into the accuracy and efficiency of the model, helping to identify how well it predicts future sales compared to actual values.

#### Comparison of Actual vs. Predicted Sales:

By plotting and comparing **Actual vs. Predicted sales**, the model demonstrated reasonable accuracy. The predicted sales values were close to the actual sales data, which shows that the model effectively captures the underlying patterns in the historical sales data. While there were small deviations, the overall trend followed by the predicted values was consistent with the actual sales data.

This comparison not only highlights the ability of the model to predict future sales with accuracy but also underscores the potential of further refining the model through advanced techniques or the inclusion of additional features to improve predictive performance.

**Actual vs Predicted Sales Visualization**  
  


# Conclusion

The **Linear Regression model** developed in this project provides valuable and accurate sales forecasts for Adidas US by analyzing historical sales data. By using key features such as Price per Unit, Units Sold, and Operating Profit, the model effectively captures the relationship between these variables and overall sales performance. This allows Adidas US to make informed decisions and implement strategies that align with predicted market trends and consumer demand.

**Summary of Benefits:**

1. **Inventory Management Optimization:**
   * Accurate sales predictions help maintain the right balance of stock, reducing stockouts and overstocking.
   * Preemptive inventory adjustments ensure a smoother supply chain.
2. **Improved Sales Strategies:**
   * Sales forecasts enable setting accurate sales targets.
   * Focus on profitable product categories and regions with strong sales growth patterns.
   * Effective pricing strategies increase revenue while maintaining customer satisfaction.
3. **Financial Planning and Budgeting:**
   * Accurate sales forecasts support effective budgeting and resource allocation.
   * Forecasting future sales helps predict cash flow and adjust marketing budgets.

**Potential Future Improvements:**

1. **Incorporating Additional Features:**
   * Seasonal trends, customer demographics, and external factors can enhance the model's accuracy.
2. **Testing Advanced Machine Learning Models:**
   * Random Forest, XGBoost, and Neural Networks can capture complex patterns and improve predictive accuracy.
3. **Cross-Validation and Model Tuning:**
   * Fine-tuning through cross-validation and hyperparameter optimization can improve accuracy and generalization.
4. **Ensemble Methods:**
   * Combining multiple models (e.g., Linear Regression with Random Forest or XGBoost) can create a more robust ensemble model.

#### Business Impact:

By leveraging data-driven insights from the Linear Regression model, Adidas US is better equipped to align its sales, inventory, and financial strategies with predicted future demand. The model not only provides actionable insights into current sales trends but also forms the foundation for more advanced forecasting techniques. With the addition of new features or advanced models, Adidas US can continue to refine its sales predictions, ensuring that the company stays ahead of market changes and consumer preferences.

As the company evolves, having a scalable forecasting framework that adapts to more complex business environments will be crucial for sustaining growth and maintaining competitive advantages in the retail industry.

# Literature

<https://scikit-learn.org/1.5/modules/generated/sklearn.linear_model.LinearRegression.html>

<https://www.geeksforgeeks.org/ml-linear-regression/>

<https://blog.hubspot.com/sales/regression-analysis-to-forecast-sales>

<https://www.kaggle.com/datasets/heemalichaudhari/adidas-sales-dataset>

[https://github.com/Ismat-Samadov/Sales\_Forecasting](https://www.kaggle.com/datasets/heemalichaudhari/adidas-sales-dataset)

<https://sales-forecasting-wiun.onrender.com/>