NATIONAL INSTITUTE OF ELECTRONICS AND INFORMATION TECHNOLOGY

# PROJECT REPORT ON

# WALMART SALES PREDICTION USING MACHINE LEARNING


**Under the Guidance of**

**Dr. Sarwan Singh**

**Ms. Anita Budhiraja**

**Ms. Ravneet Kaur**

**Submitted By:**

**Golugula Akshaya Prabha**

**1878772**

**Academic Session: 2025-2026**

# INDEX

# ACKNOWLEDGEMENT

This project has been completed with the support, guidance, and encouragement of several individuals who contributed directly or indirectly to its success. I would like to express my sincere gratitude to my project guide and faculty members for their continuous guidance, constructive feedback, and encouragement throughout the development of this project.

I would like to express my sincere appreciation to all those who provided guidance, support, and encouragement throughout the successful completion of this **Walmart Sales Demand Forecasting** project. The development of this machine learning application would not have been possible without their valuable contributions.

I am deeply grateful to **Dr. Sarwan Singh**, **Ms. Anita Budhiraja**, and **Ms. Ravneet Kaur** for their continuous mentorship, insightful suggestions, and expert guidance throughout the course of this project. Their knowledge of machine learning concepts and data analysis techniques played a crucial role in shaping the overall implementation.

I also extend my heartfelt thanks to the faculty members and technical staff of the **National Institute of Electronics and Information Technology (NIELIT), Ropar**, for providing the necessary infrastructure, resources, and academic support required to complete this work. Special appreciation is given to the open-source community for offering powerful tools and libraries such as scikit-learn, pandas, and Flask, which greatly facilitated the development of this project.

Finally, I would like to thank my family and peers for their constant motivation, encouragement, and support throughout this academic journey.

**Thank you.**

**__Golugula Akshaya Prabha**

# DECLARATION

I hereby declare that the project entitled **“Walmart Sales Demand Forecasting Using Machine Learning”**, submitted in partial fulfillment of the requirements of the **Fundamentals of Artificial Intelligence Technology** course at the **National Institute of Electronics and Information Technology (NIELIT), Ropar**, is an original work carried out by me under the guidance of **Dr. Sarwan Singh**, **Ms. Anita Budhiraja**, and **Ms. Ravneet Kaur**.

The project focuses on developing a machine learning-based prediction system to forecast weekly sales demand for Walmart departments using historical sales data. All datasets, tools, libraries, and reference materials used in this project have been appropriately acknowledged.

I further declare that this project has not been submitted previously to any institution for the award of any degree or diploma.

**Submitted by:**
**Golugula Akshaya Prabha**
**1878772**

# GUIDE CERTIFICATE

This is to certify that the project entitled **"Walmart Sales Demand Forecasting Using Machine Learning"** is a bonafide work carried out by **G Akshaya Prabha**, Registration Number: **1878772**, in partial fulfillment of the requirements for the **Fundamentals of Artificial Intelligence Technology** course at the **National Institute of Electronics and Information Technology (NIELIT), Ropar**.

It is an authentic work carried out by her at NIELIT Ropar under the guidance of **Dr. Sarwan Singh, Ms. Anita Budhiraja,** and **Ms. Ravneet Kaur.**

The project has been completed under our guidance and supervision during the academic session **2025–2026**. Throughout the project duration, the student has demonstrated a good understanding of machine learning concepts, data preprocessing techniques, and practical implementation skills.

We hereby recommend this project for evaluation.

**Project Guide**
(Signature)

# CENTER CERTIFICATE

This is to certify that **G Akshaya Prabha**, Registration Number: **1878772,** has successfully completed the project entitled **“Walmart Sales Demand Forecasting Using Machine Learning”** as part of the **Fundamentals of Artificial Intelligence Technology** course at the **National Institute of Electronics and Information Technology (NIELIT), Ropar**.

The project has been evaluated and found satisfactory in fulfilling the academic requirements of the course during the academic session **2025–2026**.

The student has demonstrated a sound understanding of machine learning concepts, data preprocessing methods, and predictive model development techniques. The project work reflects sincere effort, technical competence, and dedication throughout its execution.

**Project Guides:**
**Dr. Sarwan Singh**
**Ms. Anita Budhiraja**
**Ms. Ravneet Kaur**

Signature: _____
Date: _____

# INTRODUCTION TO THE PROJECT

In today’s competitive retail environment, accurate sales forecasting plays a vital role in effective business planning and decision-making. Retail companies rely on precise sales predictions to manage inventory levels efficiently, minimize overstocking and stock shortages, reduce operational costs, and enhance overall customer satisfaction. Reliable demand forecasting also supports better supply chain coordination and strategic planning.

This project focuses on predicting future weekly sales for Walmart departments using machine learning techniques. By analyzing historical sales data, the system learns underlying patterns, seasonal trends, and sales variations over time. The trained model forecasts upcoming sales demand, enabling informed decision-making for inventory and resource management. The project integrates data preprocessing, predictive modeling, and web-based deployment using Python and Flask to develop a complete, practical, and user-friendly sales forecasting application.

# NEED FOR STRUCTURED DATASETS

In a machine learning–based **Walmart Sales Demand Forecasting** system, structured datasets play a crucial role in ensuring accurate, efficient, and reliable predictions. Unlike unstructured data such as raw text or images, structured datasets are organized in a well-defined tabular format, which enables effective data preprocessing, model training, and evaluation.

In this project, the structured dataset includes important fields such as:

- Store Number
- Department ID
- Date
- Weekly Sales
- Holiday Indicator

This structured organization allows the machine learning model to learn meaningful patterns and trends from historical sales data, leading to more accurate demand forecasts. It also supports:

- Scalability across multiple stores and departments
- Faster data processing and model training
- Easy generation of reports and sales visualizations

Structured datasets also simplify the detection of data inconsistencies such as missing values, duplicate records, or abnormal sales figures. Moreover, they enable continuous improvement of the forecasting model through retraining with updated data.

Therefore, creating and maintaining well-structured datasets is essential for successful sales demand forecasting and requires careful data curation, validation, and collaboration between data analysts and system developers to ensure high data quality throughout the project lifecycle.

# OBJECTIVE OF THE PROJECT

The primary aim of this project is to develop a machine learning-based system capable of predicting future weekly sales **using historical Walmart sales data**. The specific objectives include:

- Understanding and analyzing retail sales data
- Applying machine learning regression techniques
- Predicting weekly sales for selected departments
- Developing a simple and user-friendly web interface
- Visualizing sales trends for better interpretation

# SCOPE OF THE PROJECT

This report presents a comprehensive overview of the design and implementation of a **Walmart Sales Demand Forecasting System** using machine learning techniques. The report covers the following key aspects:

- Collection and analysis of historical Walmart sales data.
- Dataset structuring and preprocessing for machine learning.
- Application of regression-based algorithms for sales prediction.
- Backend development using Flask for handling user inputs and predictions.
- Frontend interface designed using HTML to enable user interaction.
- Data visualization using Python libraries for sales trend analysis.
- Storage and handling of sales data using CSV files.

The report also discusses system limitations, potential enhancements, and future scope, including scalability and integration with real-time retail decision-making systems.

**Context: Challenges in Traditional Sales Forecasting**

In many retail organizations, sales forecasting is still performed using manual analysis or basic statistical methods, which:

- Requires significant time and effort.
- Is prone to human error and subjective judgment.
- Lacks scalability for large datasets and multiple product categories.

A machine learning–based sales demand forecasting system addresses these challenges by automating the prediction process using historical data. Such a system provides faster, more consistent, and data-driven forecasts, making it especially useful for large retail chains like Walmart, where accurate demand prediction is critical for inventory management and business planning.

# ROLE OF MACHINE LEARNING IN SALES PREDICTION

Artificial Intelligence, particularly **Machine Learning and Predictive Analytics**, plays a central role in this project. The system leverages data-driven techniques to analyze historical sales patterns and forecast future demand. The key AI components used in this project include:

- Data preprocessing and feature extraction from historical sales data.
- Application of regression-based machine learning algorithms for sales prediction.
- Model training and evaluation to learn relationships between time-based features and sales values.
- Prediction generation based on trained models for future sales demand.

This AI-driven approach enables automated and consistent sales forecasting while reducing dependency on manual analysis. By learning from historical trends, the model minimizes prediction errors and adapts to variations in sales patterns, making it effective for retail demand forecasting and decision support.

## Structured Datasets Used in the System

The structured datasets used in this sales demand forecasting system consist of well-organized historical sales records that enable efficient data analysis and prediction. The key components of the dataset include:

- **Store information:** Store number and related identifiers.
- **Department details:** Department ID representing product categories.
- **Time-related data:** Date of sales records, later converted into week numbers.
- **Sales data:** Weekly sales values for each department.
- **Holiday indicator:** Flag to denote whether a particular week is a holiday period.

This structured format ensures that the machine learning model can accurately associate sales values with corresponding time and department information. It also allows the system to efficiently filter, analyze, and predict sales demand for specific departments.

## Data Preprocessing Techniques

Data preprocessing is an essential step to improve the accuracy and reliability of the sales forecasting model. The preprocessing steps used in this project include:

- Converting date fields into a standard datetime format.
- Extracting week numbers from date values to create numerical features.
- Filtering the dataset based on selected department IDs.
- Handling missing or inconsistent records to maintain data quality.
- Structuring the dataset into input and output variables suitable for machine learning models.

These preprocessing steps ensure that clean and meaningful data is provided to the machine learning algorithm, resulting in more accurate and consistent predictions.

## Algorithm Used

The sales demand forecasting system primarily uses the following algorithm:

● **Linear Regression:** A regression-based machine learning algorithm used to predict continuous numerical values such as weekly sales demand.

The Linear Regression model learns the relationship between the independent variable (week number) and the dependent variable (weekly sales). Although more complex algorithms exist, Linear Regression is chosen for its simplicity, interpretability, and suitability for academic projects.

Unlike deep learning approaches, this project does not involve training complex neural networks. Instead, it focuses on applying a supervised machine learning regression model to generate reliable forecasts based on historical sales data.

# SYSTEM ARCHITECTURE

The system architecture defines the complete workflow of the application. The major steps involved are:

1. User enters department ID and future week
2. Flask application receives and processes the input
3. Dataset is filtered and preprocessed
4. Machine learning model is trained on historical data
5. Prediction is generated for the selected week
6. Predicted value and sales trend graph are displaye

## User Interface Considerations

Although the system interface is kept simple, it is designed to support effective interaction and ease of use. The user interface includes:

- A web-based frontend developed using HTML and Flask.
- Input forms for selecting department IDs and future weeks.
- A results page displaying predicted sales values.
- Graphical visualization of historical and predicted sales trends rendered through the Flask backend.

The primary objective of the interface design is to ensure clarity, accessibility, and usability for users such as analysts, managers, and decision-makers, without requiring advanced technical knowledge.

# Ethical Considerations in Sales Forecasting Systems

Machine learning–based sales forecasting systems must address important ethical considerations related to data usage, transparency, and responsible decision-making. Although this project does not involve personal data, ethical practices remain essential. Key considerations include:

- **Data Integrity:** Ensuring that sales data used for training is accurate and free from manipulation.
- **Transparency:** Clearly explaining how predictions are generated and how the model operates.
- **Responsible Usage:** Ensuring predictions are used as decision-support tools rather than absolute outcomes.
- **Fairness:** Avoiding biased interpretations that could negatively impact specific departments or business units.

To ensure ethical compliance, the project uses publicly available or authorized datasets, avoids sensitive personal information, and clearly communicates the limitations of predictions.

## Integration with Retail Management Systems

Although the current system is implemented as a standalone web application using Flask, it can be integrated into larger retail management or enterprise systems. Such integration would enable:

- Automatic synchronization with inventory management systems.
- Integration with reporting dashboards and analytics tools.
- Real-time demand forecasting for decision-makers.

Future enhancements may include the use of REST APIs, database integration, or cloud-based services to connect sales forecasts with inventory planning, supply chain optimization, and business intelligence platforms.

## Evaluation Metrics for Performance

The performance of the Walmart sales demand forecasting system is evaluated using the following metrics:

- **Prediction Accuracy:** Measures how closely the predicted sales values match the actual historical sales data.
- **Error Metrics:** Mean Absolute Error (MAE) and Mean Squared Error (MSE) are used to quantify prediction errors.
- **Model Consistency:** Evaluates how stable the predictions are across different time periods.
- **Processing Time:** Time taken to train the model and generate sales predictions.
- **User Feedback:** Assesses ease of use, clarity of outputs, and usefulness of visualizations.

For example, a lower MAE value indicates that the predicted sales are closer to the actual sales figures, reflecting better model performance.

## Limitations of the Current System

Although the system performs effectively, it has certain limitations:
● **Limited Feature Set:** Predictions are primarily based on historical sales and time-based features, without considering external factors such as promotions or weather.
● **Algorithm Simplicity:** Linear Regression may not capture complex or nonlinear sales patterns.
● **Data Dependency:** The accuracy of predictions heavily depends on the quality and quantity of historical data.
● **Static Model:** The model does not automatically update itself with new data in real time.

These limitations can be addressed by incorporating additional features, advanced algorithms, and automated model retraining.

Although the system provides useful predictions, it has certain limitations:

- Prediction accuracy depends heavily on historical data
- Linear Regression cannot capture complex non-linear patterns
- External factors such as promotions and economic conditions are not considered

## Future Directions in Sales Forecasting Development

To further enhance the system, future improvements may include:

- **Advanced Algorithms:** Implementation of models such as Random Forest, XGBoost, or LSTM for improved accuracy.
- **Real-Time Data Integration:** Automatic updates using live sales data streams.
- **Inventory Optimization:** Linking demand forecasts with inventory management systems.
- **Cloud Deployment:** Scalable cloud-based forecasting and analytics solutions.
- **Business Intelligence Integration:** Interactive dashboards for deeper insights and reporting.

Integrating external factors such as promotions, holidays, and regional demand patterns can further improve the accuracy and business relevance of the forecasting system.

## Case Studies / Real-World Applications

Sales demand forecasting systems are widely used across the retail industry to support data-driven decision-making. Some real-world applications include:

- **Large Retail Chains (e.g., Walmart):** Forecasting product demand to optimize inventory and reduce stockouts.
- **E-commerce Platforms:** Predicting future sales trends to plan promotions and pricing strategies.
- **Supply Chain Management:** Anticipating demand to improve logistics and warehouse planning.

These systems have demonstrated improved inventory efficiency, reduced operational costs, and enhanced customer satisfaction through better product availability and planning.

## Stakeholder Involvement

The development and deployment of a sales demand forecasting system involves multiple stakeholders, including:

- **Software Developers:** Responsible for implementing the backend logic, data handling, and web interface.
- **Data Scientists / ML Engineers:** Design, train, and optimize machine learning models for accurate predictions.
- **Business Managers:** Use forecast results for inventory planning and strategic decisions.
- **Retail Analysts:** Provide domain knowledge and feedback on model outputs and usability.

Collaboration among these stakeholders ensures that the system is technically robust and aligned with real-world business requirements.

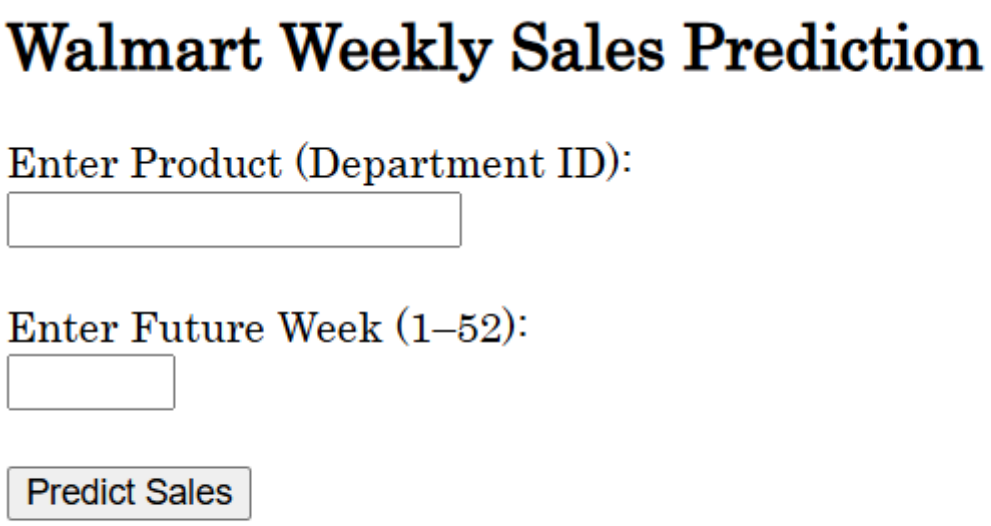
## Regulatory and Compliance Aspects

Although this project is developed for academic purposes, real-world sales forecasting systems must adhere to regulatory and compliance standards. Key considerations include:

- **Data Governance Policies:** Ensuring proper handling and authorization of business data.
- **Transparency Requirements:** Clearly documenting how forecasts are generated and used.
- **Security Measures:** Protecting sales data from unauthorized access through access control and encryption.

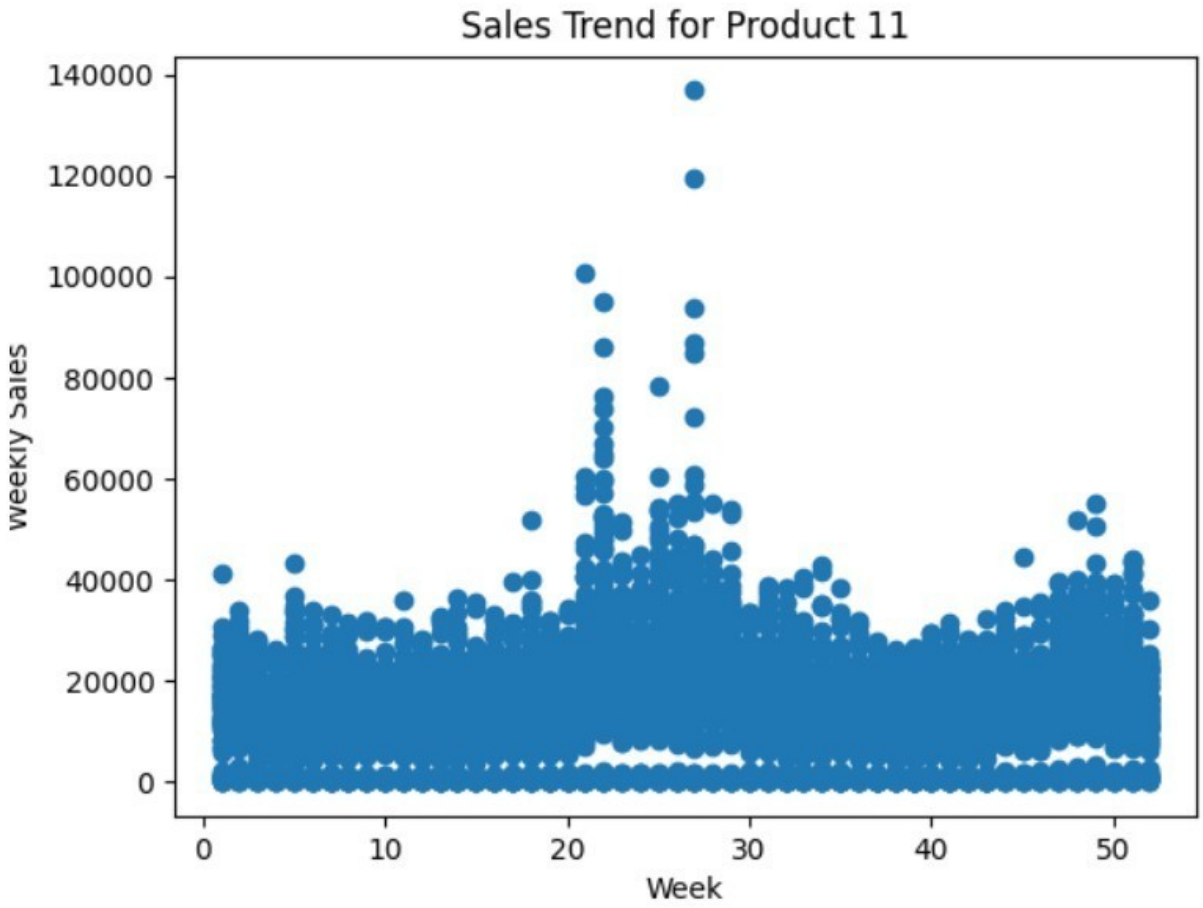
Failure to comply with data governance and security standards can lead to operational risks and loss of business trust, especially when forecasting systems are deployed at scale.

**WEB PAGE:**


Walmart Weekly Sales Prediction
Enter Product (Department ID):
Enter Future Week (1–52):
Predict Sales

## OUTPUT:

**Predicted Weekly Sales: 13853.15708423179**



# Hardware and Software Requirements

**To ensure smooth development, execution, and deployment of the Walmart Sales Demand Forecasting System, the following hardware and software specifications are recommended.**

### 1. Hardware Requirements

**This project is lightweight and does not require high-end hardware. The system can run efficiently on basic computing devices.**

#### Minimum Hardware Requirements

- **Processor: Intel Pentium IV or equivalent**
- **RAM: 2 GB**
- **Storage: 500 MB free space (for storing datasets, model files, and logs)**
- **Display: 1024 × 768 resolution**
- **Input Devices: Keyboard, Mouse**
- **Internet: Basic internet connectivity (for deployment or updates)**

#### Recommended Hardware Requirements

- **Processor: Intel Core i3 or higher**
- **RAM: 4 GB or more**
- **Storage: 1 GB or more**
- **Display: 1366 × 768 or higher resolution**
- **Network: Stable broadband internet connection**

### 2. Software Requirements

**The project is developed using Python and Flask as core technologies, supported by data analysis and machine learning libraries.**

#### Operating System

- **Windows 10 / 11**
- **Linux (Ubuntu, Debian)**
- **macOS**

#### Programming Language

- **Python 3.x (Python 3.8 or later is recommended)**

#### Required Python Libraries

- **flask – for building the web-based user interface**
- **pandas – for data loading, cleaning, and manipulation**

- **numpy – for numerical computations**
- **scikit-learn – for implementing machine learning algorithms**
- **matplotlib – for data visualization and plotting graphs**
- **os – for file and directory handling**

**3. Code Editor / IDE**

- **Visual Studio Code**
- **PyCharm**
- **Any Python-compatible code editor**

**4. Web Browser**

- **Google Chrome**
- **Mozilla Firefox**
- **Microsoft Edge**
- **Any modern web browser**

**5. Optional Tools**

- **Jupyter Notebook – for data exploration and model testing**
- **Git – for version control**
- **Excel – for viewing or editing CSV sales datasets**

**6. Installation Guide (Windows Example)**

**To set up the project environment on a Windows system, follow these steps:**

1. **Install Python from the official website:** **https://www.python.org**
2. **Open Command Prompt and install the required Python libraries:**

```
pip install flask pandas numpy scikit-learn matplotlib
```

3. **Place all project files (app.py, dataset CSV file, templates folder) in a single directory.**
4. **Run the Flask application using the command:**

```
python app.py
```

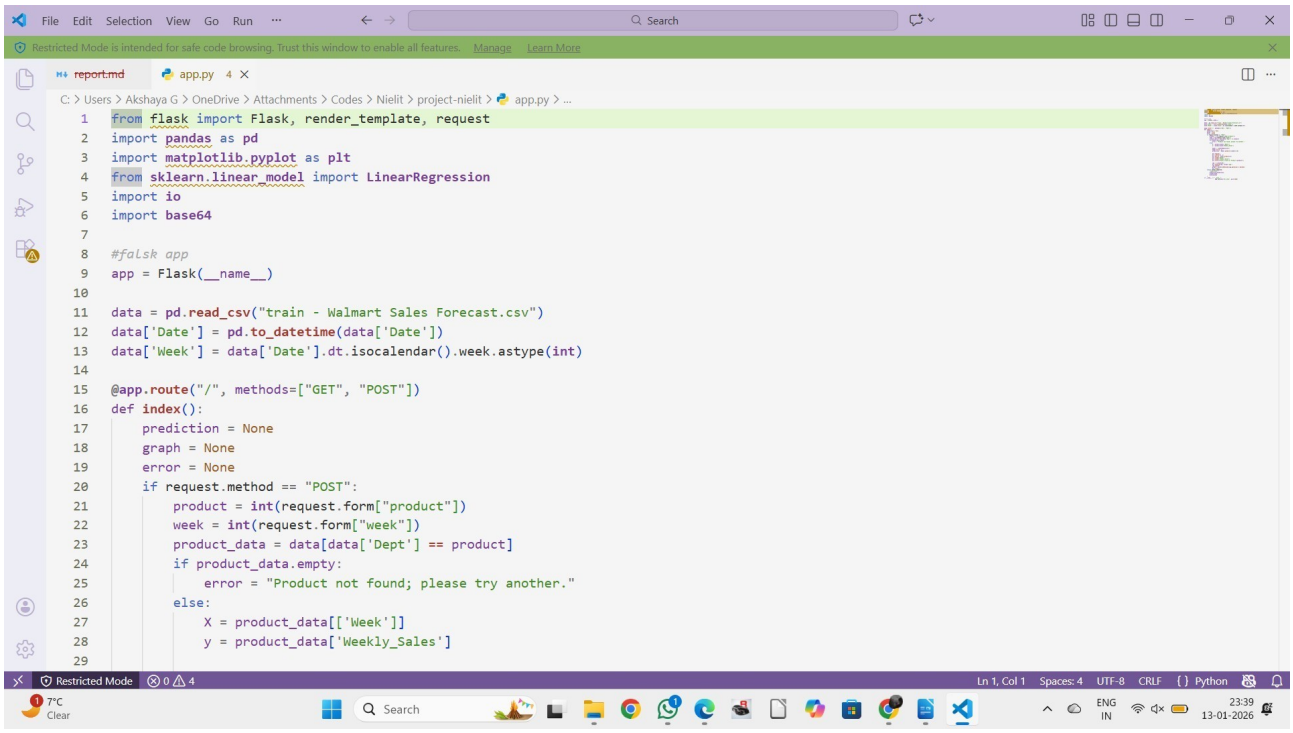
5. **Open a web browser and access the application at:**

**http://localhost:5000/**

## 7. Summary Table

| Component | Requirement |
|---|---|
| Processor | Pentium IV / Core i3 or higher |
| RAM | Minimum 2 GB (4 GB recommended) |
| Storage | 500 MB – 1 GB |
| Operating System | Windows / Linux / macOS |
| Python Version | Python 3.x |
| Libraries | flask, pandas, numpy, scikit-learn, matplotlib |
| Editor | VS Code / PyCharm |
| Browser | Chrome / Firefox |
| Internet | Basic to stable broadband |

# SOURCE CODE

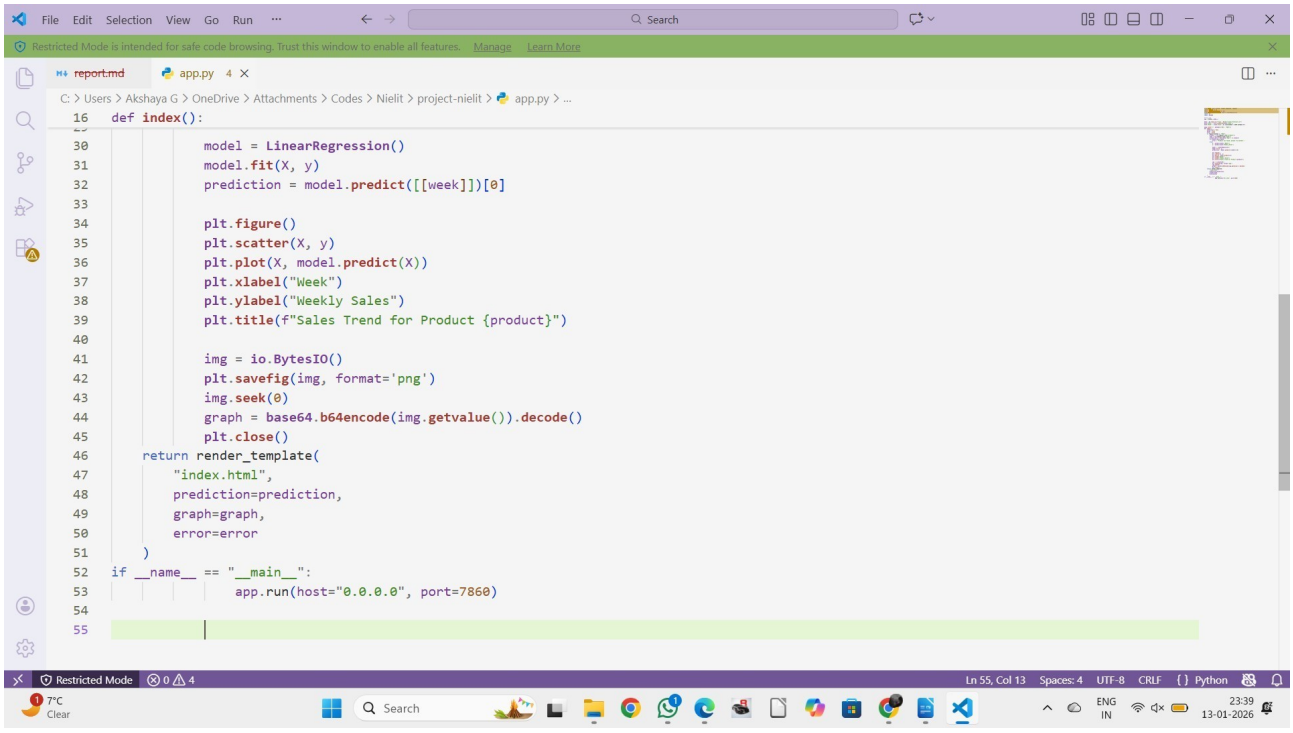


```python
from flask import Flask, render_template, request
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
import io
import base64

#falsk app
app = Flask(__name__)

data = pd.read_csv("train - Walmart Sales Forecast.csv")
data['Date'] = pd.to_datetime(data['Date'])
data['Week'] = data['Date'].dt.isocalendar().week.astype(int)

@app.route("/", methods=["GET", "POST"])
def index():
    prediction = None
    graph = None
    error = None
    if request.method == "POST":
        product = int(request.form["product"])
        week = int(request.form["week"])
        product_data = data[data['Dept'] == product]
        if product_data.empty:
            error = "Product not found; please try another."
        else:
            X = product_data[['Week']]
            y = product_data['Weekly_Sales']
```



```python
def index():
            model = LinearRegression()
            model.fit(X, y)
            prediction = model.predict([[week]])[0]

            plt.figure()
            plt.scatter(X, y)
            plt.plot(X, model.predict(X))
            plt.xlabel("Week")
            plt.ylabel("Weekly Sales")
            plt.title(f"Sales Trend for Product {product}")

            img = io.BytesIO()
            plt.savefig(img, format='png')
            img.seek(0)
            graph = base64.b64encode(img.getvalue()).decode()
            plt.close()
    return render_template(
        "index.html",
        prediction=prediction,
        graph=graph,
        error=error
    )
if __name__ == "__main__":
            app.run(host="0.0.0.0", port=7860)
```

# Conclusion

**The Walmart Sales Demand Forecasting System developed using Python and machine learning provides an effective and practical solution to the challenges of manual and traditional sales prediction methods. The system helps retailers make data-driven decisions by accurately forecasting future weekly sales demand, thereby improving inventory planning and reducing operational inefficiencies.**

**With features such as automated sales prediction, structured data handling using CSV files, graphical visualization of sales trends, and a user-friendly web interface built with Flask, the project establishes a strong foundation for intelligent retail analytics. The integration of machine learning techniques demonstrates how historical data can be effectively leveraged to support business decision-making.**

**In the future, this system can be enhanced by incorporating advanced forecasting algorithms, real-time data integration, cloud deployment, and database support. Such improvements would make the application more scalable, robust, and suitable for large-scale retail environments in the modern digital era.**