Sales Forecasting ML Application

A Streamlit web application that allows users to upload pre-processed CSV files and automatically applies machine learning models to generate sales statistics and visualizations.

Features

- **Interactive Dashboard**: Clean and intuitive web interface
- **Automated ML**: Automatically trains and evaluates multiple ML models
- **Rich Visualizations**: Generates comprehensive sales analysis charts
- **Export Results**: Download summary statistics and model performance metrics
- **Configurable**: Adjustable parameters for model training

Required CSV Format

Your uploaded CSV file should contain the following columns:

- **Product Name** - Name of the product
- **Total Sales** - Sales amount
- **Date** - Transaction date (YYYY-MM-DD format)
- **Temperature** - Temperature on that day
- **Holiday(0/1)** - Holiday indicator (0=No, 1=Yes)
- **Day** - Day of week (1=Monday, 7=Sunday)

Installation

1. Install the required packages:

```
pip install -r requirements.txt
```

1. Run the application:

```
streamlit run app.py
```

1. Open your browser and navigate to `http://localhost:8501`

How to Use

1. **Upload Data**: Click on the file uploader and select your pre-processed CSV file
2. **Review Overview**: Check the data overview and sample data display
3. **Configure Models**: Adjust the minimum samples and maximum products settings
4. **Train Models**: Click "Train Models" to automatically train ML models for your products
5. **View Results**: Review model performance metrics and feature importance
6. **Generate Visualizations**: Click "Generate Visualizations" to create sales analysis charts
7. **Download Results**: Export summary statistics for further analysis

Visualizations Generated

The application creates the following visualizations similar to your reference images:

- **Top 10 Products by Sales**: Horizontal bar chart showing best-performing products
- **Temperature vs Sales**: Scatter plot showing relationship between temperature and sales
- **Holiday vs Non-Holiday Sales**: Bar chart comparing sales on holidays vs regular days
- **Sales by Day of Week**: Bar chart showing sales patterns across weekdays

Machine Learning Models

The application automatically selects the best performing model from:

- **Random Forest Regressor**: Ensemble method good for non-linear relationships
- **Gradient Boosting Regressor**: Advanced ensemble method for complex patterns
- **Linear Regression**: Simple baseline model for linear relationships

Model Features

The ML models use the following engineered features:

- Sales lag features (1, 2, 3, 7 days)
- Rolling average features (3, 7 days)
- Cyclical day of week features (sin/cos encoding)
- Cyclical month features (sin/cos encoding)
- Temperature
- Holiday indicator

Performance Metrics

For each trained model, the application provides:

- **MSE (Mean Squared Error)**: Lower values indicate better performance
- **MAE (Mean Absolute Error)**: Average prediction error in original units
- **R^2 (R-squared)**: Proportion of variance explained (higher is better)

Technical Details

- Built with Streamlit for the web interface
- Uses scikit-learn for machine learning models
- Matplotlib and Seaborn for visualizations
- Pandas for data manipulation
- Time series cross-validation for model evaluation

Notes

- Products with insufficient data (less than minimum samples) will be skipped
- The application uses time series cross-validation to prevent data leakage
- Feature importance is calculated to understand which factors drive sales
- All visualizations are interactive and can be downloaded

Troubleshooting

If you encounter issues:

1. Ensure your CSV file has all required columns
2. Check that date formats are consistent (YYYY-MM-DD)
3. Verify that numeric columns contain valid numbers
4. Make sure you have sufficient data points per product

For best results, ensure you have at least 10-20 data points per product you want to model.