# LEVEL 2:

# Multiple Linear Regression – Boston Housing Dataset

## Objective

The objective of this task was to build a multiple linear regression model to predict median house prices (MEDV) using socio-economic and environmental variables from the Boston Housing dataset.

## Data Overview

- Total observations: 506
- Total variables: 14
- Target variable: MEDV (Median House Value)
- No missing values
- No duplicate rows detected

The dataset was split into:

- 80% Training Data
- 20% Testing Data

## Model Performance

### Training Model Results

- Multiple R-squared: **0.7441**
- Adjusted R-squared: **0.7355**
- Residual Standard Error: 4.731
- F-statistic: 87.21 ($p < 0.001$)

This indicates that approximately **74.41 percent** of the variance in housing prices is explained by the independent variables.

### Test Set Evaluation

- Mean Squared Error (MSE): **23.68**
- Test R-squared: **0.719**

The model explains approximately **71.9 percent** of unseen test data variation, indicating good generalization performance.

# 4. Key Significant Predictors

Statistically significant variables include:

- CRIM (Crime rate) – Negative impact
- ZN (Residential land proportion) – Positive impact
- CHAS (Near river) – Positive impact
- NOX (Pollution level) – Strong negative impact
- RM (Number of rooms) – Strong positive impact
- DIS (Distance to employment centers) – Negative impact

- RAD (Highway accessibility) – Positive impact
- TAX – Negative impact
- PTRATIO (Pupil-teacher ratio) – Negative impact
- B (Proportion of Black population variable) – Positive impact
- LSTAT (Lower income percentage) – Strong negative impact

### Most Influential Variables

- RM: Each additional room increases house price significantly.
- LSTAT: Higher lower-income population reduces house prices.
- NOX: Higher pollution significantly decreases housing value.

## Interpretation

The regression model demonstrates strong explanatory power and reasonable predictive accuracy. Socioeconomic factors and environmental quality significantly influence housing prices. The model performs consistently on both training and testing data, indicating minimal overfitting.

# Time Series Analysis – Apple Stock Prices

## 1. Objective

The objective of this task was to analyze Apple Inc. closing stock prices using time series visualization, moving averages, and decomposition techniques.

## 2. Time Series Visualization

The Apple stock closing price plot shows:

- Clear upward long-term trend from 2014 to 2018
- Periodic short-term fluctuations
- Increased volatility in later years

This indicates steady growth with cyclical variations.

## 30-Day Moving Average

A 30-day moving average was calculated to smooth short-term fluctuations.

### Observations:

- The moving average closely follows the underlying trend.
- Short-term noise is reduced.
- Trend direction becomes clearer.
- Strong bullish pattern observed toward 2017–2018.

The moving average confirms sustained upward momentum in the stock price.

## Time Series Decomposition

The decomposition separated the series into:

- Trend Component
- Seasonal Component
- Random Component

### Trend

The trend shows consistent long-term growth.

### Seasonal

Minimal strong seasonal pattern is observed, which is expected in daily stock data.

### Random (Irregular)

Irregular fluctuations represent market volatility, news impacts, and investor behavior.

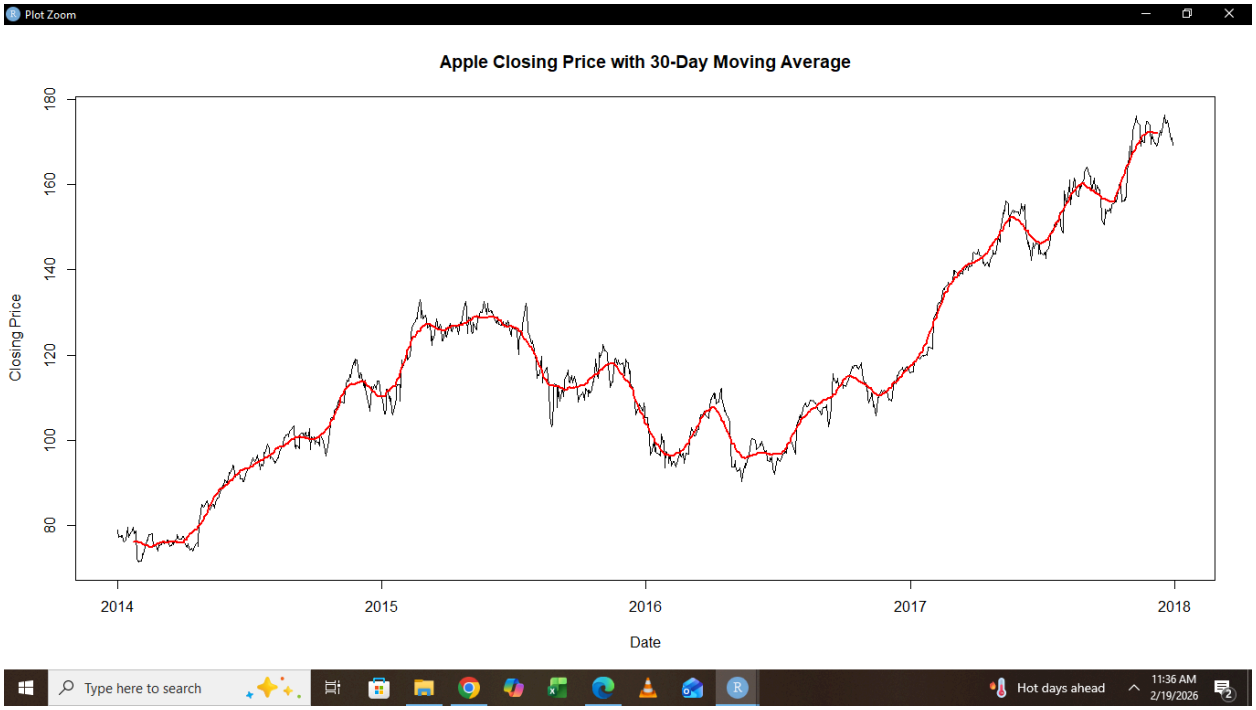
## 5. Interpretation

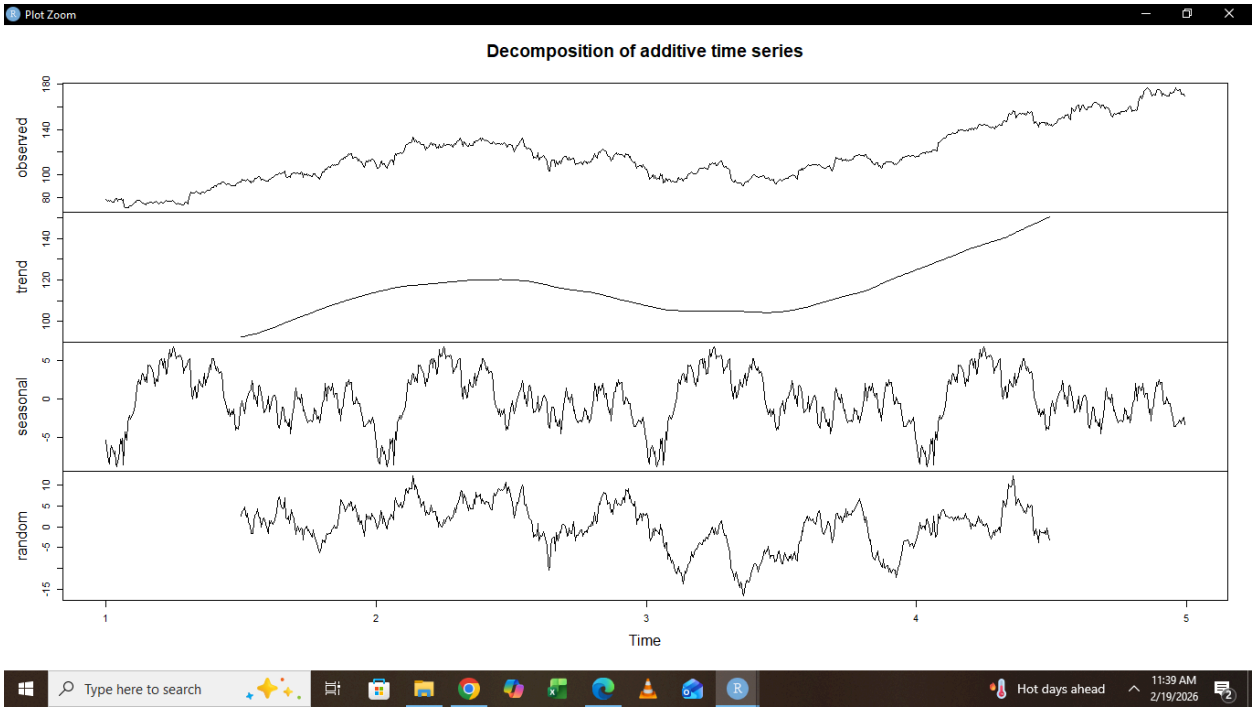
The Apple stock exhibits:

- Strong long-term growth trend
- Moderate volatility
- Limited clear seasonal structure
- Significant random market-driven movements

The moving average effectively highlights the underlying growth trajectory.

## Visualisations


Plot Zoom
Apple Closing Price with 30-Day Moving Average
Closing Price
80
100
120
140
160
180
2014
2015
2016
2017
2018
Date
Type here to search
Hot days ahead
11:36 AM
2/19/2026


Plot Zoom
Decomposition of additive time series
observed
trend
seasonal
random
Time
Type here to search
Hot days ahead
11:39 AM
2/19/2026

# Overall Conclusion

Across both tasks:

- Regression modeling demonstrated strong predictive capability in structured economic data.

- Time series analysis demonstrated ability to identify trends, smooth volatility, and decompose market-driven data.