ML USING WATSON

PROJECT

BY

JEFFREY THOMAS R

ESWAR.G

TABLE OF CONTENT

S no	Content	Page no
1.	INTRODUCTION	3
2.	R PROGRAM	4
3.	LINEAR REGRESSION	5-6
4.	IMPORTING DATA	7
5.	IMPLEMENTING LINEAR REGRESSION IN R	8
6.	PROGRAM	9
7.	OUTPUT	10-24
8.	CONCLUSION	25

INTRODUCTION

1.1. Overview of Stock Price Prediction

Stock price prediction has long been a subject of interest for investors, traders, and financial analysts. It involves forecasting the future prices of a company's stock based on historical data and various predictive models. Accurate stock price prediction can lead to profitable investment decisions and effective risk management strategies. With the advent of sophisticated analytical tools and computational techniques, predicting stock prices has become more precise and accessible.

1.2. Importance of Stock Price Prediction

The ability to predict stock prices is crucial for multiple reasons:

- **Investment Decisions**: Investors rely on price predictions to make informed decisions about buying, holding, or selling stocks.
- **Risk Management**: Predicting potential price movements helps in assessing risks and mitigating losses.
- **Market Efficiency**: Accurate predictions contribute to market efficiency by reflecting all available information in the stock prices.
- **Economic Implications**: Stock prices often reflect broader economic conditions, making predictions valuable for economic forecasting.

1.3. Objectives of the Document

This document aims to:

- Provide an in-depth understanding of linear regression as a predictive model.
- Demonstrate the application of linear regression in the context of stock price prediction using the R programming language.
- Guide readers through the process of data collection, preprocessing, model implementation, and evaluation.
- Present a case study to illustrate the practical application and challenges of stock price prediction.
- Offer insights into improving model accuracy and addressing potential limitations.

2. R program

2.1. Introduction to R

R is a powerful, open-source programming language and software environment designed specifically for statistical computing and graphics. Developed in the early 1990s by Ross Ihaka and Robert Gentleman at the University of Auckland, R has grown to become a leading tool for data analysis, visualization, and statistical modeling. Its extensive range of statistical techniques, combined with the ability to produce high-quality graphical representations, makes R a preferred choice for data scientists, statisticians, and researchers worldwide.

2.2. Why Use R for Stock Price Prediction?

R offers several advantages that make it particularly well-suited for stock price prediction:

- **Comprehensive Statistical Analysis**: R provides a vast array of built-in functions for statistical modeling, which are essential for analyzing and predicting stock prices.
- **Data Manipulation and Cleaning**: With packages like `dplyr` and `tidyr`, R excels in data manipulation and preprocessing, which are crucial steps in building accurate predictive models.
- **Visualization Capabilities**: R's powerful visualization libraries, such as `ggplot2`, allow for detailed and insightful graphical representation of data, aiding in the understanding and interpretation of stock price trends.
- **Extensive Libraries**: R boasts a rich ecosystem of packages specifically designed for financial analysis, including `quantmod`, `TTR`, and `PerformanceAnalytics`, which simplify the process of retrieving, analyzing, and modeling financial data.
- **Community Support**: The active and growing R community contributes to an ever-expanding pool of resources, tutorials, and forums, making it easier for users to find solutions and share knowledge.

2.3. Key Packages and Libraries in R for Linear Regression

To effectively use R for stock price prediction using linear regression, several key packages are essential:

- **`ggplot2`**: A powerful package for creating advanced and customizable visualizations, making it easier to understand and present data.
- **`dplyr`**: Facilitates efficient data manipulation, allowing for streamlined data cleaning and preparation.
- **`caret`**: Provides a unified interface for training and evaluating machine learning models, including linear regression.
- **`quantmod`**: Specializes in quantitative financial modeling, making it easy to retrieve and manage stock market data.
- **`TTR`**: Offers technical trading rules, making it valuable for generating technical indicators that can enhance predictive models.
- **`PerformanceAnalytics`**: Useful for performance and risk analysis of financial instruments and portfolios.

3.linear regression

3.1. What is Linear Regression?

Linear regression is a fundamental statistical method used for modeling the relationship between a dependent variable and one or more independent variables. The core idea is to establish a linear equation that best fits the observed data points. The equation takes the form:

$Y=\beta0+\beta1X1+\beta2X2+\ldots+\betanXn+\epsilonY = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_nX_n + \epsilonY=\beta0+\beta1X1+\beta2X2+\ldots+\betanXn+\epsilon$

where:

- YYY is the dependent variable (the variable we are trying to predict),
- $\beta0\beta_0\beta0$ is the y-intercept of the regression line,
- $\beta1,\beta2,\ldots,\betan\beta_1, \beta_2, \ldots, \beta_n\beta1,\beta2,\ldots,\betan$ are the coefficients that represent the impact of each independent variable ($X1,X2,\ldots,XnX_1, X_2, \ldots, X_nX1,X2,\ldots,Xn$) on the dependent variable,
- $\epsilon\epsilonϵ$ is the error term (the difference between the observed and predicted values).

The primary goal of linear regression is to determine the values of the coefficients ($\beta0,\beta1,\ldots,\betan\beta_0, \beta_1, \ldots, \beta_n\beta0,\beta1,\ldots,\betan$) that minimize the sum of squared residuals (the differences between observed and predicted values).

3.2. Types of Linear Regression

- **Simple Linear Regression**: This involves a single independent variable and aims to establish a linear relationship between the dependent and independent variable.

  Example: Predicting stock prices based on historical prices.

  $Y=\beta0+\beta1X+\epsilonY = \beta_0 + \beta_1X + \epsilonY=\beta0+\beta1X+\epsilon$

- **Multiple Linear Regression**: This extends simple linear regression by incorporating two or more independent variables to predict the dependent variable.

Example: Predicting stock prices based on multiple factors such as historical prices, trading volume, and market indices.

$$Y=\beta0+\beta1X1+\beta2X2+\ldots+\betanXn+\epsilonY = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_nX_n + \epsilonY=\beta0+\beta1X1+\beta2X2+\ldots+\betanXn+\epsilon$$

- **Polynomial Regression**: This is a form of multiple linear regression where the independent variable is raised to a power to model non-linear relationships.

  Example: Predicting stock prices where the relationship between the price and the influencing factors is non-linear.

  $$Y=\beta0+\beta1X+\beta2X2+\ldots+\betanXn+\epsilonY = \beta_0 + \beta_1X + \beta_2X^2 + \ldots + \beta_nX^n + \epsilonY=\beta0+\beta1X+\beta2X2+\ldots+\betanXn+\epsilon$$

3.3. Applications of Linear Regression in Finance

Linear regression is widely used in finance due to its simplicity and effectiveness in modeling relationships between variables. Some common applications include:

- **Stock Price Prediction**: By analyzing historical stock prices and other financial indicators, linear regression can help forecast future stock prices.
- **Risk Management**: Identifying and quantifying the impact of various risk factors on financial instruments and portfolios.
- **Economic Forecasting**: Predicting economic indicators such as GDP growth, inflation rates, and unemployment rates based on historical data and other economic variables.
- **Portfolio Optimization**: Assessing the relationship between asset returns and risk factors to optimize investment portfolios.

4. importing data

One of the crucial steps in building a stock price prediction model is importing data into R. Various packages in R can be used to import data from different sources and formats. This section will focus on the `readr` package and other relevant packages for handling different data formats like CSV, Excel, and more.

Using `readr` and Other Packages

- **`readr` Package:**

The `readr` package provides a fast and friendly way to read rectangular data. It is part of the tidyverse collection of packages, which are designed for data science.

- **Installing `readr`**:

```
install.packages("readr")

library(readr)
```

- **Reading CSV Files**:

```
data <- read_csv("path/to/your/file.csv")
```

The `read_csv` function is used to read CSV files into R. It automatically determines the column types and handles missing values.

5.Implemention linear regression in r

Visualizing the Data

Effective data visualization is a crucial step in understanding the underlying patterns and relationships in your data. In the context of stock price prediction, visualizing stock prices and trends, as well as performing correlation analysis using visual tools, can provide valuable insights.

Plotting Stock Prices and Trends

To plot stock prices and visualize trends over time, we can use the `ggplot2` package, which offers a powerful and flexible system for creating high-quality plots.

- **Installing and Loading `ggplot2`**:

```
install.packages("ggplot2")
library(ggplot2)
```

- **Plotting Stock Prices**:

```
ggplot(data, aes(x = Date, y = Close)) +
  geom_line(color = "blue") +
  labs(title = "Stock Price Over Time", x
= "Date", y = "Closing Price") +
  theme_minimal()
```

6.program

```
install.packages("readr")
install.packages("dplyr")
install.packages("ggplot2")
library(readr)
library(dplyr)
library(ggplot2)
# Replace 'your_data.csv' with your actual file path
stock_data <- read.csv(file.choose("C:\\Users\\admin\\Desktop\\IBM.csv"))
names(stock_data)
ggplot(stock_data, aes(date, close,group =TRUE)) +
  geom_line(color = "blue") +
  labs(
    title = "Stock Price",
    subtitle = "Source: Yahoo Finance",
    y = "Stock Price (in Dollars)"
  )
# Fit the model
linearmodel=lm(close~date, data = stock_data)
# View model summary
summary(linearmodel)
# Replace 'specific_date' with the desired date
specific_date <- as.Date("2024-06-22")  # Example specific date
# Create a data frame with the specific date
new_data <- data.frame(Date = specific_date)
# Predict using the model
predicted_price <- predict(linearmodel, newdata = stock_data)
# Print the predicted price
print(predicted_price)
```

7.Output

```
library(ggplot2)
> library(dplyr)
> # Replace 'your_data.csv' with your actual file path
> stock_data <- read.csv(file.choose("C:\\Users\\admin\\Desktop\\IBM.csv"))
> names(stock_data)
 [1] "date"             "Open"             "High"
 [4] "Low"              "close"            "Volume"
 [7] "SD20"             "Upper_Band"       "Lower_Band"
[10] "S_Close.t.1."     "S_Close.t.2."     "S_Close.t.3."
[13] "S_Close.t.5."     "S_Open.t.1."      "MA5"
[16] "MA10"             "MA20"             "MA50"
[19] "MA200"            "EMA10"            "EMA20"
[22] "EMA50"            "EMA100"           "EMA200"
[25] "MACD"             "MACD_EMA"         "ATR"
[28] "ADX"              "CCI"              "ROC"
[31] "RSI"              "William.R"        "SO.K"
[34] "STD5"             "ForceIndex1"      "ForceIndex20"
[37] "Date_col"         "Day"              "DayofWeek"
[40] "DayofYear"        "Week"             "Is_month_end"
[43] "Is_month_start"   "Is_quarter_end"   "Is_quarter_start"
[46] "Is_year_end"      "Is_year_start"    "Is_leap_year"
[49] "Year"             "Month"            "QQQ_Close"
[52] "QQQ.t.1."         "QQQ.t.2."         "QQQ.t.5."
[55] "QQQ_MA10"         "QQQ_MA20"         "QQQ_MA50"
[58] "SnP_Close"        "SnP.t.1.."        "SnP.t.5."
[61] "DJIA_Close"       "DJIA.t.1.."       "DJIA.t.5."
[64] "Close_forcast"
> ggplot(stock_data, aes(date, close,group =TRUE)) +
+  geom_line(color = "blue") +
+  labs(
+    title = "Stock Price",
```

```
+    subtitle = "Source: kaggle",
+    y = "Stock Price (in Dollars)"
+  )
> # Fit the model
> linearmodel=lm(close~date, data = stock_data)
> # View model summary
> summary(linearmodel)

Call:
lm(formula = close ~ date, data = stock_data)

Residuals:
ALL 4988 residuals are 0: no residual degrees of freedom!

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     59.61        NaN     NaN      NaN
date1/10/02     18.69        NaN     NaN      NaN
date1/10/03     -2.99        NaN     NaN      NaN
date1/10/05      3.13        NaN     NaN      NaN
date1/10/06     -3.97        NaN     NaN      NaN
date1/10/07      6.72        NaN     NaN      NaN
date1/10/08      8.36        NaN     NaN      NaN
date1/10/11     46.61        NaN     NaN      NaN
date1/10/12     73.04        NaN     NaN      NaN
date1/10/13     83.88        NaN     NaN      NaN
date1/10/14     82.40        NaN     NaN      NaN
date1/10/17     80.61        NaN     NaN      NaN
date1/10/18     84.87        NaN     NaN      NaN
date1/10/19     52.42        NaN     NaN      NaN
date1/10/20     72.15        NaN     NaN      NaN
date1/11/01      0.16        NaN     NaN      NaN
date1/11/02     17.52        NaN     NaN      NaN
date1/11/05      2.68        NaN     NaN      NaN
```

date1/11/06	-3.90	NaN	NaN	NaN
date1/11/07	6.56	NaN	NaN	NaN
date1/11/08	6.83	NaN	NaN	NaN
date1/11/10	31.79	NaN	NaN	NaN
date1/11/11	46.35	NaN	NaN	NaN
date1/11/12	73.78	NaN	NaN	NaN
date1/11/13	85.05	NaN	NaN	NaN
date1/11/16	48.08	NaN	NaN	NaN
date1/11/17	82.50	NaN	NaN	NaN
date1/11/18	84.89	NaN	NaN	NaN
date1/11/19	52.11	NaN	NaN	NaN
date1/12/01	0.24	NaN	NaN	NaN
date1/12/04	-0.04	NaN	NaN	NaN
date1/12/05	2.82	NaN	NaN	NaN
date1/12/06	-4.30	NaN	NaN	NaN
date1/12/07	7.02	NaN	NaN	NaN
date1/12/09	-0.29	NaN	NaN	NaN
date1/12/10	32.52	NaN	NaN	NaN
date1/12/11	47.66	NaN	NaN	NaN
date1/12/12	72.48	NaN	NaN	NaN
date1/12/15	61.91	NaN	NaN	NaN
date1/12/16	47.81	NaN	NaN	NaN
date1/12/17	82.67	NaN	NaN	NaN
date1/12/18	83.96	NaN	NaN	NaN
date1/13/03	-3.10	NaN	NaN	NaN
date1/13/04	-1.24	NaN	NaN	NaN
date1/13/05	2.32	NaN	NaN	NaN
date1/13/06	-4.56	NaN	NaN	NaN
date1/13/09	-0.55	NaN	NaN	NaN
date1/13/10	32.32	NaN	NaN	NaN
date1/13/11	47.46	NaN	NaN	NaN
date1/13/12	71.46	NaN	NaN	NaN
date1/13/14	80.05	NaN	NaN	NaN
date1/13/15	62.19	NaN	NaN	NaN

date1/13/16	46.42	NaN	NaN	NaN
date1/13/17	82.15	NaN	NaN	NaN
date1/13/20	72.07	NaN	NaN	NaN
date1/14/02	16.07	NaN	NaN	NaN
date1/14/03	-2.41	NaN	NaN	NaN
date1/14/04	-0.85	NaN	NaN	NaN
date1/14/05	2.09	NaN	NaN	NaN
date1/14/08	10.40	NaN	NaN	NaN
date1/14/09	-2.03	NaN	NaN	NaN
date1/14/10	33.79	NaN	NaN	NaN
date1/14/11	48.31	NaN	NaN	NaN
date1/14/13	83.68	NaN	NaN	NaN
date1/14/14	81.38	NaN	NaN	NaN
date1/14/15	61.41	NaN	NaN	NaN
date1/14/16	47.82	NaN	NaN	NaN
date1/14/19	51.13	NaN	NaN	NaN
date1/14/20	71.31	NaN	NaN	NaN
date1/15/02	16.58	NaN	NaN	NaN
date1/15/03	-3.05	NaN	NaN	NaN
date1/15/04	1.57	NaN	NaN	NaN
date1/15/08	9.66	NaN	NaN	NaN
date1/15/09	-1.39	NaN	NaN	NaN
date1/15/10	33.42	NaN	NaN	NaN
date1/15/13	83.59	NaN	NaN	NaN
date1/15/14	82.76	NaN	NaN	NaN
date1/15/15	60.45	NaN	NaN	NaN
date1/15/16	45.49	NaN	NaN	NaN
date1/15/19	52.36	NaN	NaN	NaN
date1/15/20	72.08	NaN	NaN	NaN
date1/16/01	-0.44	NaN	NaN	NaN
date1/16/02	15.65	NaN	NaN	NaN
date1/16/03	-4.04	NaN	NaN	NaN
date1/16/04	2.41	NaN	NaN	NaN
date1/16/07	8.02	NaN	NaN	NaN

date1/16/08	9.52	NaN	NaN	NaN
date1/16/09	-0.84	NaN	NaN	NaN
date1/16/13	83.66	NaN	NaN	NaN
date1/16/14	83.54	NaN	NaN	NaN
date1/16/15	62.45	NaN	NaN	NaN
date1/16/18	84.58	NaN	NaN	NaN
date1/16/19	52.26	NaN	NaN	NaN
date1/16/20	73.40	NaN	NaN	NaN
date1/17/01	2.07	NaN	NaN	NaN
date1/17/02	17.25	NaN	NaN	NaN
date1/17/03	-7.11	NaN	NaN	NaN
date1/17/06	-4.67	NaN	NaN	NaN
date1/17/07	7.48	NaN	NaN	NaN
date1/17/08	9.16	NaN	NaN	NaN
date1/17/12	72.08	NaN	NaN	NaN
date1/17/13	84.45	NaN	NaN	NaN
date1/17/14	84.55	NaN	NaN	NaN
date1/17/17	82.62	NaN	NaN	NaN
date1/17/18	88.81	NaN	NaN	NaN
date1/17/19	52.78	NaN	NaN	NaN
date1/17/20	73.71	NaN	NaN	NaN
date1/18/01	9.49	NaN	NaN	NaN
date1/18/02	13.63	NaN	NaN	NaN
date1/18/05	2.62	NaN	NaN	NaN
date1/18/06	-4.14	NaN	NaN	NaN
date1/18/07	7.10	NaN	NaN	NaN
date1/18/08	10.72	NaN	NaN	NaN
date1/18/11	48.77	NaN	NaN	NaN
date1/18/12	72.86	NaN	NaN	NaN
date1/18/13	85.06	NaN	NaN	NaN
date1/18/17	81.70	NaN	NaN	NaN
date1/18/18	89.22	NaN	NaN	NaN
date1/18/19	54.28	NaN	NaN	NaN
date1/19/01	11.36	NaN	NaN	NaN

date1/19/05	1.44	NaN	NaN	NaN
date1/19/06	-4.61	NaN	NaN	NaN
date1/19/07	4.90	NaN	NaN	NaN
date1/19/10	35.08	NaN	NaN	NaN
date1/19/11	52.40	NaN	NaN	NaN
date1/19/12	72.46	NaN	NaN	NaN
date1/19/16	43.94	NaN	NaN	NaN
date1/19/17	81.71	NaN	NaN	NaN
date1/19/18	83.28	NaN	NaN	NaN
date1/2/01	-5.51	NaN	NaN	NaN
date1/2/02	18.28	NaN	NaN	NaN
date1/2/03	-7.58	NaN	NaN	NaN
date1/2/04	-0.04	NaN	NaN	NaN
date1/2/08	11.60	NaN	NaN	NaN
date1/2/09	0.86	NaN	NaN	NaN
date1/2/13	86.46	NaN	NaN	NaN
date1/2/14	81.09	NaN	NaN	NaN
date1/2/15	66.27	NaN	NaN	NaN
date1/2/18	76.13	NaN	NaN	NaN
date1/2/19	46.36	NaN	NaN	NaN
date1/2/20	70.93	NaN	NaN	NaN
date1/20/04	3.57	NaN	NaN	NaN
date1/20/05	1.37	NaN	NaN	NaN
date1/20/06	-5.76	NaN	NaN	NaN
date1/20/09	-2.87	NaN	NaN	NaN
date1/20/10	32.34	NaN	NaN	NaN
date1/20/11	52.48	NaN	NaN	NaN
date1/20/12	78.31	NaN	NaN	NaN
date1/20/15	62.30	NaN	NaN	NaN
date1/20/16	38.89	NaN	NaN	NaN
date1/20/17	84.87	NaN	NaN	NaN
date1/21/03	-7.60	NaN	NaN	NaN
date1/21/04	3.96	NaN	NaN	NaN
date1/21/05	0.97	NaN	NaN	NaN

date1/21/09	3.66	NaN	NaN	NaN
date1/21/10	31.45	NaN	NaN	NaN
date1/21/11	52.26	NaN	NaN	NaN
date1/21/14	83.29	NaN	NaN	NaN
date1/21/15	58.53	NaN	NaN	NaN
date1/21/16	39.74	NaN	NaN	NaN
date1/21/20	74.54	NaN	NaN	NaN
date1/22/01	9.64	NaN	NaN	NaN
date1/22/02	11.23	NaN	NaN	NaN
date1/22/03	-8.14	NaN	NaN	NaN
date1/22/04	3.84	NaN	NaN	NaN
date1/22/07	5.53	NaN	NaN	NaN
date1/22/08	9.24	NaN	NaN	NaN
date1/22/09	2.73	NaN	NaN	NaN
date1/22/10	28.98	NaN	NaN	NaN
date1/22/13	86.26	NaN	NaN	NaN
date1/22/14	78.60	NaN	NaN	NaN
date1/22/15	61.09	NaN	NaN	NaN
date1/22/16	39.41	NaN	NaN	NaN
date1/22/18	83.48	NaN	NaN	NaN
date1/22/19	53.09	NaN	NaN	NaN
date1/22/20	79.09	NaN	NaN	NaN
date1/23/01	9.96	NaN	NaN	NaN
date1/23/02	9.56	NaN	NaN	NaN
date1/23/03	-7.27	NaN	NaN	NaN
date1/23/04	4.09	NaN	NaN	NaN
date1/23/06	-5.73	NaN	NaN	NaN
date1/23/07	5.51	NaN	NaN	NaN
date1/23/08	12.56	NaN	NaN	NaN
date1/23/09	2.33	NaN	NaN	NaN
date1/23/12	79.38	NaN	NaN	NaN
date1/23/13	92.69	NaN	NaN	NaN
date1/23/14	78.96	NaN	NaN	NaN
date1/23/15	61.46	NaN	NaN	NaN

date1/23/17	85.28	NaN	NaN	NaN
date1/23/18	86.69	NaN	NaN	NaN
date1/23/19	62.63	NaN	NaN	NaN
date1/23/20	78.11	NaN	NaN	NaN
date1/24/01	10.84	NaN	NaN	NaN
date1/24/02	10.09	NaN	NaN	NaN
date1/24/03	-8.60	NaN	NaN	NaN
date1/24/05	0.58	NaN	NaN	NaN
date1/24/06	-6.10	NaN	NaN	NaN
date1/24/07	5.72	NaN	NaN	NaN
date1/24/08	13.11	NaN	NaN	NaN
date1/24/11	55.23	NaN	NaN	NaN
date1/24/12	80.81	NaN	NaN	NaN
date1/24/13	92.46	NaN	NaN	NaN
date1/24/14	76.62	NaN	NaN	NaN
date1/24/17	89.41	NaN	NaN	NaN
date1/24/18	85.92	NaN	NaN	NaN
date1/24/19	62.30	NaN	NaN	NaN
date1/24/20	75.88	NaN	NaN	NaN
date1/25/01	11.04	NaN	NaN	NaN
date1/25/02	10.45	NaN	NaN	NaN
date1/25/05	0.84	NaN	NaN	NaN
date1/25/06	-6.06	NaN	NaN	NaN
date1/25/07	5.80	NaN	NaN	NaN
date1/25/08	11.48	NaN	NaN	NaN
date1/25/10	29.42	NaN	NaN	NaN
date1/25/11	56.54	NaN	NaN	NaN
date1/25/12	80.66	NaN	NaN	NaN
date1/25/13	92.87	NaN	NaN	NaN
date1/25/16	39.07	NaN	NaN	NaN
date1/25/17	91.43	NaN	NaN	NaN
date1/25/18	86.01	NaN	NaN	NaN
date1/25/19	63.62	NaN	NaN	NaN
date1/26/01	13.23	NaN	NaN	NaN

```
date1/26/04      5.36        NaN     NaN      NaN
date1/26/05      0.68        NaN     NaN      NaN
date1/26/06     -6.18        NaN     NaN      NaN
date1/26/07      5.76        NaN     NaN      NaN
date1/26/09      3.79        NaN     NaN      NaN
date1/26/10     29.16        NaN     NaN      NaN
date1/26/11     56.25        NaN     NaN      NaN
date1/26/12     80.11        NaN     NaN      NaN
date1/26/15     61.85        NaN     NaN      NaN
date1/26/16     39.48        NaN     NaN      NaN
date1/26/17     91.74        NaN     NaN      NaN
date1/26/18     87.65        NaN     NaN      NaN
date1/27/03     -8.97        NaN     NaN      NaN
date1/27/04      4.68        NaN     NaN      NaN
date1/27/05      0.70        NaN     NaN      NaN
date1/27/06     -5.98        NaN     NaN      NaN
date1/27/09      3.83        NaN     NaN      NaN
date1/27/10     29.57        NaN     NaN      NaN
date1/27/11     56.27        NaN     NaN      NaN
date1/27/12     79.73        NaN     NaN      NaN
date1/27/14     75.30        NaN     NaN      NaN
date1/27/15     59.76        NaN     NaN      NaN
date1/27/16     38.16        NaN     NaN      NaN
date1/27/17     90.59        NaN     NaN      NaN
date1/27/20     74.01        NaN     NaN      NaN
date1/28/02      9.72        NaN     NaN      NaN
date1/28/03     -7.88        NaN     NaN      NaN
date1/28/04      3.75        NaN     NaN      NaN
 [ reached getOption("max.print") -- omitted 4738 rows ]

Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared:      1,   Adjusted R-squared:    NaN
F-statistic:   NaN on 4987 and 0 DF,  p-value: NA
```

```
> specific_date <- as.Date("2024-06-22")  # Example specific date
> # Create a data frame with the specific date
> new_data <- data.frame(Date = specific_date)
> # Predict using the model
> predicted_price <- predict(linearmodel, newdata = stock_data)
> # Print the predicted price
> print(predicted_price)
    1     2     3     4     5     6     7     8     9    10    11    12
70.80 71.99 60.80 61.44 60.37 59.17 58.26 55.79 59.09 59.69 59.45 62.76
   13    14    15    16    17    18    19    20    21    22    23    24
62.80 64.95 63.79 63.91 65.19 63.79 63.43 59.33 62.16 63.47 63.39 62.68
   25    26    27    28    29    30    31    32    33    34    35    36
65.03 65.87 62.84 62.84 63.75 62.80 62.44 63.67 59.65 61.00 62.76 65.95
   37    38    39    40    41    42    43    44    45    46    47    48
61.72 59.41 61.88 60.60 59.89 58.21 58.97 56.02 57.73 57.49 54.86 52.03
   49    50    51    52    53    54    55    56    57    58    59    60
56.78 54.10 54.02 54.38 54.22 54.10 60.36 59.45 59.97 59.69 59.05 59.61
   61    62    63    64    65    66    67    68    69    70    71    72
59.77 59.85 59.17 61.68 69.10 70.97 69.25 69.57 70.45 70.65 72.84 73.35
   73    74    75    76    77    78    79    80    81    82    83    84
74.39 71.45 72.76 70.34 71.59 72.84 74.67 72.87 71.53 73.38 72.65 73.51
   85    86    87    88    89    90    91    92    93    94    95    96
74.58 73.45 71.21 68.66 69.55 66.42 67.25 65.52 63.80 67.73 65.33 67.00
   97    98    99   100   101   102   103   104   105   106   107   108
67.70 68.69 68.00 63.41 60.99 62.84 60.65 61.03 57.54 59.14 56.39 56.89
  109   110   111   112   113   114   115   116   117   118   119   120
56.90 59.72 60.93 63.55 60.30 60.70 61.43 60.46 57.73 58.76 62.72 62.56
  121   122   123   124   125   126   127   128   129   130   131   132
61.31 63.26 62.22 61.44 61.79 63.67 68.02 73.11 73.34 71.53 71.96 73.35
  133   134   135   136   137   138   139   140   141   142   143   144
72.64 74.21 73.53 75.69 73.70 72.62 73.99 74.02 75.26 74.80 73.66 71.49
  145   146   147   148   149   150   151   152   153   154   155   156
71.97 72.63 74.05 73.58 75.09 76.12 75.46 75.07 76.48 75.32 73.71 72.03
  157   158   159   160   161   162   163   164   165   166   167   168
```

71.49	72.18	72.66	74.79	75.13	74.97	74.24	75.04	74.97	74.52	74.01	72.64
169	170	171	172	173	174	175	176	177	178	179	180
73.07	73.43	72.31	72.00	72.17	72.03	72.28	72.59	73.60	72.57	73.12	72.24
181	182	183	184	185	186	187	188	189	190	191	192
71.68	68.10	66.96	65.20	66.40	68.58	69.40	68.94	69.40	66.68	66.50	67.59
193	194	195	196	197	198	199	200	201	202	203	204
67.68	66.83	67.07	67.78	66.95	67.68	67.27	68.46	69.57	69.17	68.11	67.94
205	206	207	208	209	210	211	212	213	214	215	216
66.71	66.64	67.20	67.78	68.00	67.23	67.71	66.97	66.65	65.24	66.56	65.95
217	218	219	220	221	222	223	224	225	226	227	228
68.50	68.42	67.20	66.67	64.26	64.00	64.98	64.25	62.75	61.84	61.77	59.76
229	230	231	232	233	234	235	236	237	238	239	240
61.72	61.47	59.80	57.94	60.70	60.47	58.46	57.62	58.73	59.36	60.04	62.07
241	242	243	244	245	246	247	248	249	250	251	252
62.30	62.76	63.07	62.20	62.27	63.64	64.56	65.31	65.21	65.88	64.83	65.72
253	254	255	256	257	258	259	260	261	262	263	264
67.39	67.75	69.51	70.85	71.17	69.55	69.57	69.19	70.36	70.11	70.41	72.74
265	266	267	268	269	270	271	272	273	274	275	276
72.98	72.96	73.13	73.63	74.81	73.31	73.56	73.40	73.72	74.00	73.30	73.95
277	278	279	280	281	282	283	284	285	286	287	288
74.57	73.21	71.89	73.36	74.10	73.16	74.77	77.82	77.02	77.18	76.71	77.89
289	290	291	292	293	294	295	296	297	298	299	300
78.98	77.09	77.63	77.79	78.34	79.42	78.66	78.21	77.86	78.47	79.17	78.79
301	302	303	304	305	306	307	308	309	310	311	312
77.54	77.89	79.27	80.52	79.52	79.94	79.81	78.30	77.13	75.68	76.19	75.26
313	314	315	316	317	318	319	320	321	322	323	324
76.86	73.24	70.84	69.17	69.70	70.06	69.33	66.03	67.66	69.16	69.23	68.47
325	326	327	328	329	330	331	332	333	334	335	336
68.14	68.45	66.70	67.39	68.93	68.41	69.37	69.26	66.05	65.03	63.75	61.87
337	338	339	340	341	342	343	344	345	346	347	348
63.20	63.10	62.36	62.80	62.98	66.13	67.98	67.83	68.23	66.57	67.46	67.55
349	350	351	352	353	354	355	356	357	358	359	360
69.65	68.80	68.43	68.55	68.27	69.00	67.72	68.54	67.79	66.48	66.05	66.37
361	362	363	364	365	366	367	368	369	370	371	372

66.76 66.03 64.80 64.16 64.73 62.43 56.11 56.32 57.14 54.04 54.95 54.79

373 374 375 376 377 378 379 380 381 382 383 384

55.33 54.44 57.10 57.13 56.44 56.06 55.52 55.66 54.38 53.85 53.77 54.11

385 386 387 388 389 390 391 392 393 394 395 396

53.83 52.50 48.78 49.11 53.03 51.41 51.25 52.86 54.98 54.35 54.96 55.11

397 398 399 400 401 402 403 404 405 406 407 408

54.32 53.67 54.03 54.25 53.45 52.79 52.48 52.90 51.74 50.24 51.01 51.81

409 410 411 412 413 414 415 416 417 418 419 420

51.23 50.36 49.40 48.55 48.01 48.62 48.99 49.61 48.84 47.18 46.04 44.22

421 422 423 424 425 426 427 428 429 430 431 432

44.83 44.12 45.05 46.24 46.31 43.48 44.11 45.35 47.27 45.86 44.81 44.22

433 434 435 436 437 438 439 440 441 442 443 444

44.64 44.51 45.66 44.38 45.47 46.34 46.31 44.06 43.12 44.73 44.60 42.71

445 446 447 448 449 450 451 452 453 454 455 456

45.78 46.17 45.28 43.90 43.66 42.44 43.67 44.59 46.16 46.30 46.26 46.35

457 458 459 460 461 462 463 464 465 466 467 468

48.29 49.31 51.15 53.17 52.39 52.21 52.85 51.83 51.19 50.25 49.03 49.39

469 470 471 472 473 474 475 476 477 478 479 480

48.59 46.64 47.53 46.53 47.18 48.02 48.73 47.83 46.33 46.73 46.62 46.25

481 482 483 484 485 sss 486 487 488 489 490 491 492

44.83 41.77 41.20 40.85 38.51 40.62 39.98 38.91 37.59 39.43 38.44 38.68

493 494 495 496 497 498 499 500 501 502 503 504

36.48 36.65 36.77 35.50 37.12 41.20 40.88 44.14 41.83 46.54 47.86 48.70

505 506 507 508 509 510 511 512 513 514 515 516

48.02 48.09 46.47 48.06 49.35 49.47 50.71 50.88 51.83 53.18 52.65 52.66

517 518 519 520 521 522 523 524 525 526 527 528

50.98 50.11 49.91 51.11 51.24 52.13 51.67 51.13 50.61 52.70 54.83 54.52

529 530 531 532 533 534 535 536 537 538 539 540

55.67 54.93 56.63 56.13 56.38 55.03 54.05 53.64 53.16 51.40 52.18 52.59

541 542 543 544 545 546 547 548 549 550 551 552

51.77 51.66 52.71 51.86 51.10 50.80 51.53 51.83 51.51 50.69 49.96 49.24

553 554 555 556 557 558 559 560 561 562 563 564

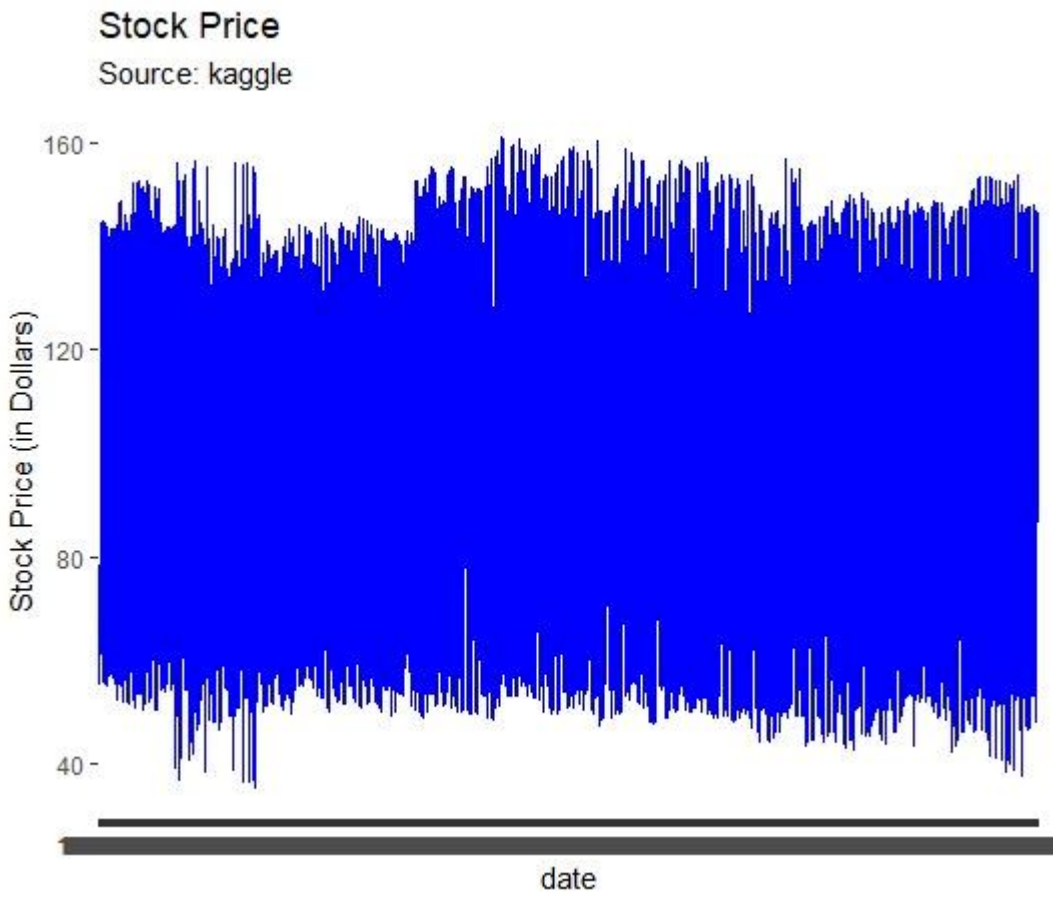
50.05 52.03 52.73 53.98 55.54 54.37 56.18 56.62 56.51 57.20 56.56 55.57

565 566 567 568 569 570 571 572 573 574 575 576

52.50	52.01	51.47	52.34	51.01	50.64	51.73	51.87	50.56	50.50	50.49	49.78
577	578	579	580	581	582	583	584	585	586	587	588
49.89	50.15	49.89	50.41	50.07	49.50	49.08	50.11	51.33	51.45	51.21	51.73
589	590	591	592	593	594	595	596	597	598	599	600
50.83	51.16	50.08	50.00	50.44	50.04	49.63	50.29	49.87	50.40	48.98	48.75
601	602	603	604	605	606	607	608	609	610	611	612
48.64	50.76	51.12	53.35	53.36	53.06	53.19	54.93	53.22	53.99	52.77	52.70
613	614	615	616	617	618	619	620	621	622	623	624
52.31	50.75	50.94	52.71	53.00	52.27	52.07	51.81	50.93	51.12	50.95	51.81
625	626	627	628	629	630	631	632	633	634	635	636
53.57	53.63	54.52	53.94	55.41	55.48	55.13	54.27	55.08	55.42	54.93	55.57
637	638	639	640	641	642	643	644	645	646	647	648
56.66	55.98	56.62	56.19	55.78	56.75	57.69	58.33	57.50	58.27	57.68	56.04
649	650	651	652	653	654	655	656	657	658	659	660
55.60	55.86	55.80	55.27	56.84	56.76	56.63	57.07	56.61	54.33	54.61	53.09
661	662	663	664	665	666	667	668	669	670	671	672
51.89	53.15	52.97	54.43	54.42	53.64	54.77	54.64	54.90	54.81	55.05	53.92
673	674	675	676	677	678	679	680	681	682	683	684
54.21	53.47	54.68	54.07	53.48	54.18	54.93	54.42	55.81	55.91	55.40	54.47
685	686	687	688	689	690	691	692	693	694	695	696
55.03	55.37	56.03	56.23	54.02	54.27	53.48	53.06	53.35	52.84	54.16	53.50
697	698	699	700	701	702	703	704	705	706	707	708
53.02	52.48	52.67	52.68	52.59	51.76	51.80	52.41	52.53	52.62	52.94	52.74
709	710	711	712	713	714	715	716	717	718	719	720
52.97	53.12	54.25	53.81	53.94	53.81	53.89	53.23	53.59	53.26	53.20	53.27
721	722	723	724	725	726	727	728	729	730	731	732
55.70	56.07	57.10	56.48	57.87	58.08	57.05	57.11	57.61	57.48	58.65	58.87
733	734	735	736	737	738	739	740	741	742	743	744
59.76	60.59	59.36	59.33	58.07	58.07	57.84	58.10	57.37	58.68	58.51	58.87
745	746	747	748	749	750	751	752	753	754	755	756
59.22	59.56	60.18	60.05	60.19	60.54	60.22	60.24	57.99	57.96	57.81	57.76
757	758	759	760	761	762	763	764	765	766	767	768
57.46	57.26	57.43	57.50	58.39	58.03	57.89	58.12	58.25	57.90	57.48	58.13
769	770	771	772	773	774	775	776	777	778	779	780

57.43	58.53	58.15	59.01	59.27	58.73	58.43	57.88	58.17	57.52	57.67	58.34
781	782	783	784	785	786	787	788	789	790	791	792
58.19	58.50	58.91	59.22	59.05	58.76	59.49	58.98	59.34	58.97	59.69	60.12
793	794	795	796	797	798	799	800	801	802	803	804
60.33	59.94	61.15	60.78	60.34	60.61	60.77	60.38	60.04	60.45	60.85	60.27
805	806	807	808	809	810	811	812	813	814	815	816
60.31	59.57	60.55	60.55	60.37	60.54	59.35	59.57	58.37	58.76	61.18	62.02
817	818	819	820	821	822	823	824	825	826	827	828
63.18	63.57	63.45	63.70	64.97	64.29	63.36	63.77	64.57	64.67	65.07	65.19
829	830	831	832	833	834	835	836	837	838	839	840
64.33	64.48	64.49	64.92	65.15	64.72	64.99	64.76	64.15	63.74	63.42	62.54
841	842	843	844	845	846	847	848	849	850	851	852
63.08	62.92	63.08	62.89	63.25	63.10	63.12	62.82	62.86	61.65	61.61	60.65
853	854	855	856	857	858	859	860	861	862	863	864
59.45	60.81	59.84	60.25	60.87	60.52	59.71	59.32	59.52	59.81	60.22	60.46
865	866	867	868	869	870	871	872	873	874	875	876
60.40	60.17	59.86	60.20	61.39	61.38	61.07	60.66	60.69	61.10	60.64	61.07
877	878	879	880	881	882	883	884	885	886	887	888
60.91	60.14	59.92	59.08	59.48	59.14	59.49	58.94	59.38	58.92	58.06	57.46
889	890	891	892	893	894	895	896	897	898	899	900
57.37	58.01	57.98	57.71	57.59	56.74	56.90	56.69	56.94	56.43	55.86	56.20
901	902	903	904	905	906	907	908	909	910	911	912
56.85	57.09	56.90	56.88	57.93	57.70	57.86	57.86	57.55	57.46	57.05	57.18
913	914	915	916	917	918	919	920	921	922	923	924
57.89	58.80	58.84	59.08	58.82	59.13	59.02	59.06	58.82	58.44	58.79	59.29
925	926	927	928	929	930	931	932	933	934	935	936
58.77	58.48	57.93	57.66	57.57	57.14	56.84	55.97	55.74	54.63	54.79	54.92
937	938	939	940	941	942	943	944	945	946	947	948
55.67	54.94	54.87	55.04	55.71	56.40	55.71	56.20	55.41	55.57	56.09	56.07
949	950	951	952	953	954	955	956	957	958	959	960
56.67	56.86	56.61	55.97	56.14	55.64	54.63	54.68	55.62	54.77	53.80	54.92
961	962	963	964	965	966	967	968	969	970	971	972
54.99	55.00	55.71	55.56	55.79	55.40	55.44	55.67	55.43	55.59	55.24	55.43
973	974	975	976	977	978	979	980	981	982	983	984

```
55.12 55.35 55.23 55.61 56.19 56.57 56.78 56.60 56.75 56.53 56.36 56.11
  985   986   987   988   989   990   991   992   993   994   995   996
56.09 56.10 55.18 54.90 55.26 55.08 55.29 55.62 56.11 56.75 57.04 57.15
  997   998   999  1000
57.62 57.21 56.75 56.70
 [ reached getOption("max.print") -- omitted 3988 entries ]
```



8.CONCLUSION

Linear regression is a powerful and widely-used statistical method for predicting continuous outcomes based on one or more predictor variables. Its simplicity, interpretability, and efficiency make it an essential tool in various fields, including finance, economics, biology, and social sciences.

Summary of Key Points

- **Understanding Linear Regression**: Linear regression establishes a relationship between a dependent variable and one or more independent variables through a linear equation. It helps in understanding the impact of each predictor on the outcome.
- **Types of Linear Regression**: The method includes simple linear regression, which involves a single predictor, and multiple linear regression, which handles multiple predictors. Polynomial regression is also a variation that models non-linear relationships.
- **Applications in Finance**: In finance, linear regression is used for predicting stock prices, assessing risk, optimizing portfolios, and forecasting economic indicators. It provides a quantitative foundation for making informed financial decisions.
- **Advantages of R Language**: The R programming language offers comprehensive libraries and packages for implementing linear regression, making it a preferred choice for data scientists and statisticians. Its capabilities in data manipulation, visualization, and statistical analysis enhance the efficiency of building and evaluating predictive models.
- **Data Visualization and Preprocessing**: Effective data visualization and preprocessing are critical steps in the modeling process. Tools like `ggplot2`, `readr`, `readxl`, and `caret` in R facilitate these tasks, ensuring that the data is clean and well-understood before modeling.
- **Model Building and Evaluation**: Building a linear regression model involves formulating the model, splitting data into training and testing sets, and training the model using functions like `lm()`. Evaluating the model with metrics such as RMSE, MAE, and R-squared helps in assessing its predictive accuracy and reliability.

Final Thoughts

Linear regression remains a fundamental technique for predictive modeling, offering a balance between simplicity and effectiveness. While it has limitations, such as the assumption of linearity and sensitivity to outliers, its application in real-world scenarios continues to provide valuable insights and reliable predictions.

For stock price prediction, linear regression serves as a robust starting point. By leveraging the powerful capabilities of the R language, analysts can build, visualize, and refine their models to achieve greater accuracy and better decision-making. As data science and machine learning continue to evolve, linear regression will remain a cornerstone method, complemented by more advanced techniques for even more sophisticated predictive analytics.