K.R. MANGALAM UNIVERSITY
THE COMPLETE WORLD OF EDUCATION

R PROGRAMMING FOR DATA SCIENCE AND DATA ANALYTICS LAB

Lab Manual

[Course Code – SEC039]

K.R. MANGALAM UNIVERSITY

BCA(AI AND DS)
Section – B
Semester II

Submitted To:

Mr. Sahil Singh Baghel

Submitted By:

ABHISHEK SINGH
2401201063

DECLARATION:

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all the principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will cause disciplinary action by the Institute and canal so evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed. We further declare that if any violation of the intellectual property right or copyright, my supervisor and university should not be held responsible for the same.

Student Name (Signature)	**Roll No.**
Mr. Sahil Singh Baghel	2401201063

Place: K.R. Mangalam University

Date:

University Vision & Mission

Vision

KR Mangalam University aspires to become an internationally recognized institution of higher learning through excellence in inter-disciplinary education, research and innovation, preparing socially responsible life-long learners contributing to nation building.

Mission

- Foster employability and entrepreneurship through futuristic curriculum and progressive pedagogy with cutting-edge technology.
- Instil notion of lifelong learning through stimulating research, Outcomes-based education, and innovative thinking.
- Integrate global needs and expectations through collaborative programs with premier universities, research centres, industries, and professional bodies.
- Enhance leadership qualities among the youth understanding ethical values and environmental realities.

School Vision & Mission

Vision

To excel in scientific and technical education through integrated teaching, research, and innovation.

Mission

- Creating a unique and innovative learning experience to enhance quality in the domain of Engineering & Technology.
- Promoting Curricular, co-curricular and extracurricular activities that support overall personality development and lifelong learning, emphasizing character building and ethical behaviour.
- Focusing on employability through research, innovation and entrepreneurial mindset development.
- Enhancing collaborations with National and International organizations and institutions to develop cross-cultural understanding to adapt and thrive in the 21st century.

CERTIFICATE

It is certified that the work contained in the project report titled **"Student Academic Performance Prediction"** by the following student:

Name of the Student:	**Roll Number:**
Abhishek singh	2401201063

has been carried out under our/my supervision and that this work has not been submitted elsewhere for a degree.

Name of the Supervisor/s: **Signature of Supervisors**

Mr. Sahil Singh Baghel

Date: - 6th May

Place: - K.R. Mangalam University

ACKNOWLEDGEMENT

“Enthusiasm is the feet of all progress, with it there is accomplishment and

Without it there are only slits alibis.”

Acknowledgment is not a ritual but is certainly an important thing for the successful completion of the project. At the time when we were made to know about the project, it was really tough to proceed further as we were to develop the same on a platform, which was new to us. More so, the coding part seemed tricky that it seemed to be impossible for us to complete the work within the given duration.

We really feel indebted in acknowledging the organizational support and encouragement received from the university.

The task of developing this system would not have been possible without the constant help of our faculty members and friends. We take this opportunity to express our profound sense of gratitude and respect to those who helped us throughout the duration of this project.

We express our gratitude to our supervisors Mr. Ashwani Kumar for giving their valuable time and guidance to us.

Place: - K.R. Mangalam University
Date: - 6th May 2025

Abhishek Singh
2401201063

INDEX

Task 1 -Time Series Forecasting

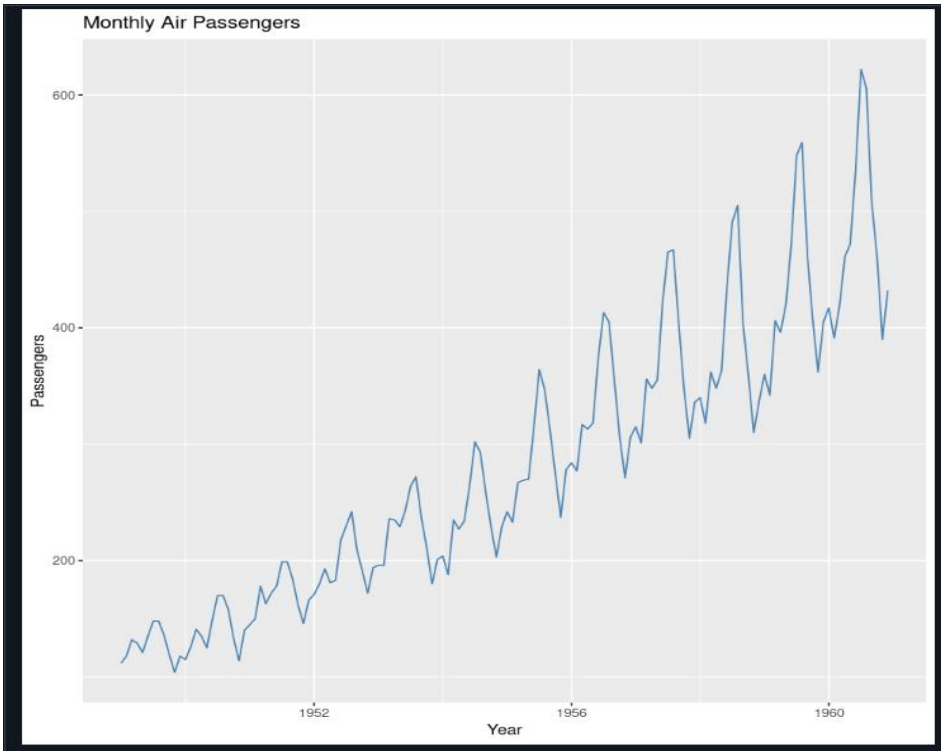
#Step 1 -Importing Libraries

```
# Install if not already installed
install.packages(c("ggplot2", "forecast", "fable", "fabletools", "tsibble", "tibble", "dplyr"))

# Load libraries
library(ggplot2)
library(forecast)
library(fable)
library(fabletools)
library(tsibble)
library(tibble)
library(dplyr)
```

```
Loading required package: fabletools
```

#Step 2 - Importing Dataset

```
data("AirPassengers")
df <- as.data.frame(AirPassengers)
df$Month <- time(AirPassengers)

# Plot using ggplot2
ggplot(df, aes(x = Month, y = AirPassengers)) +
  geom_line(color = "steelblue") +
  labs(title = "Monthly Air Passengers", x = "Year", y = "Passengers")
```




```
Don't know how to automatically pick scale for object of type <ts>. Defaulting
to continuous.
Don't know how to automatically pick scale for object of type <ts>. Defaulting
to continuous.
```

#Step 3 - Convert to Time Series Object

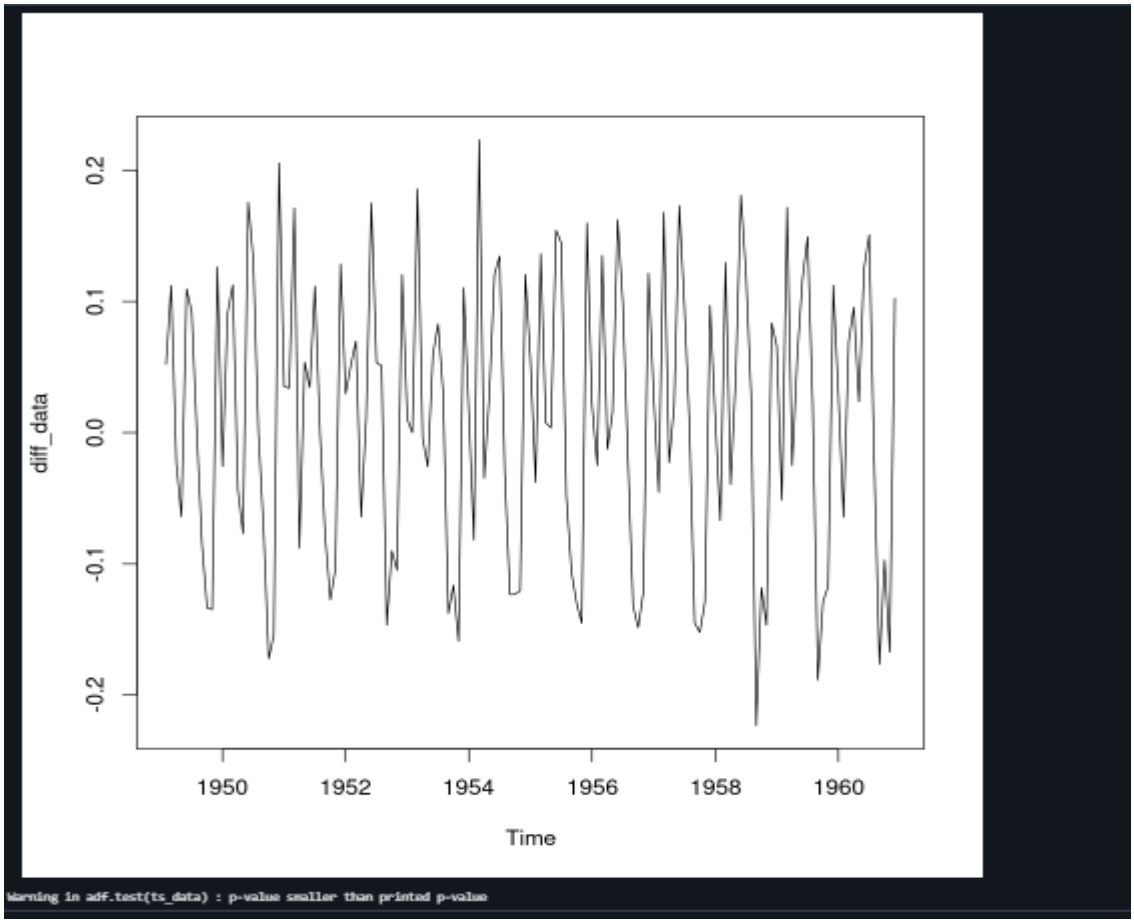
```
✓ 06:50 PM (<1s)                                                    3

# Using ts object
ts_data <- ts(AirPassengers, frequency = 12, start = c(1949, 1))

# Convert to tsibble
tsibble_data <- ts_data %>%
  as_tsibble(index = yearmonth) %>%
  rename(Passengers = value)
```

#Step 4 - Check Stationarity and Transform If Needed

```
✓ 2 minutes ago (<1s)                                               5

adf.test(ts_data)
```

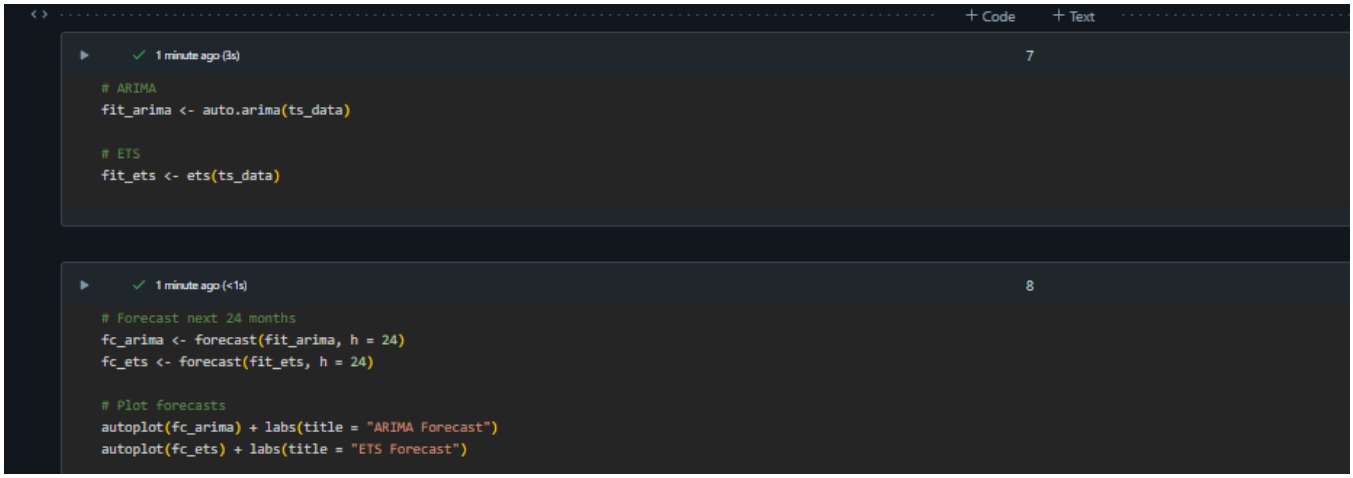
```
Warning in adf.test(ts_data) : p-value smaller than printed p-value

        Augmented Dickey-Fuller Test

data:  ts_data
Dickey-Fuller = -7.3186, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

```
✓ 1 minute ago (<1s)                                                6

# Augmented Dickey-Fuller Test
adf.test(ts_data)  # from tseries package if needed

# Visual inspection of ACF/PACF
acf(ts_data)
pacf(ts_data)

# Log transformation + differencing if non-stationary
log_data <- log(ts_data)
diff_data <- diff(log_data)
plot(diff_data)
```

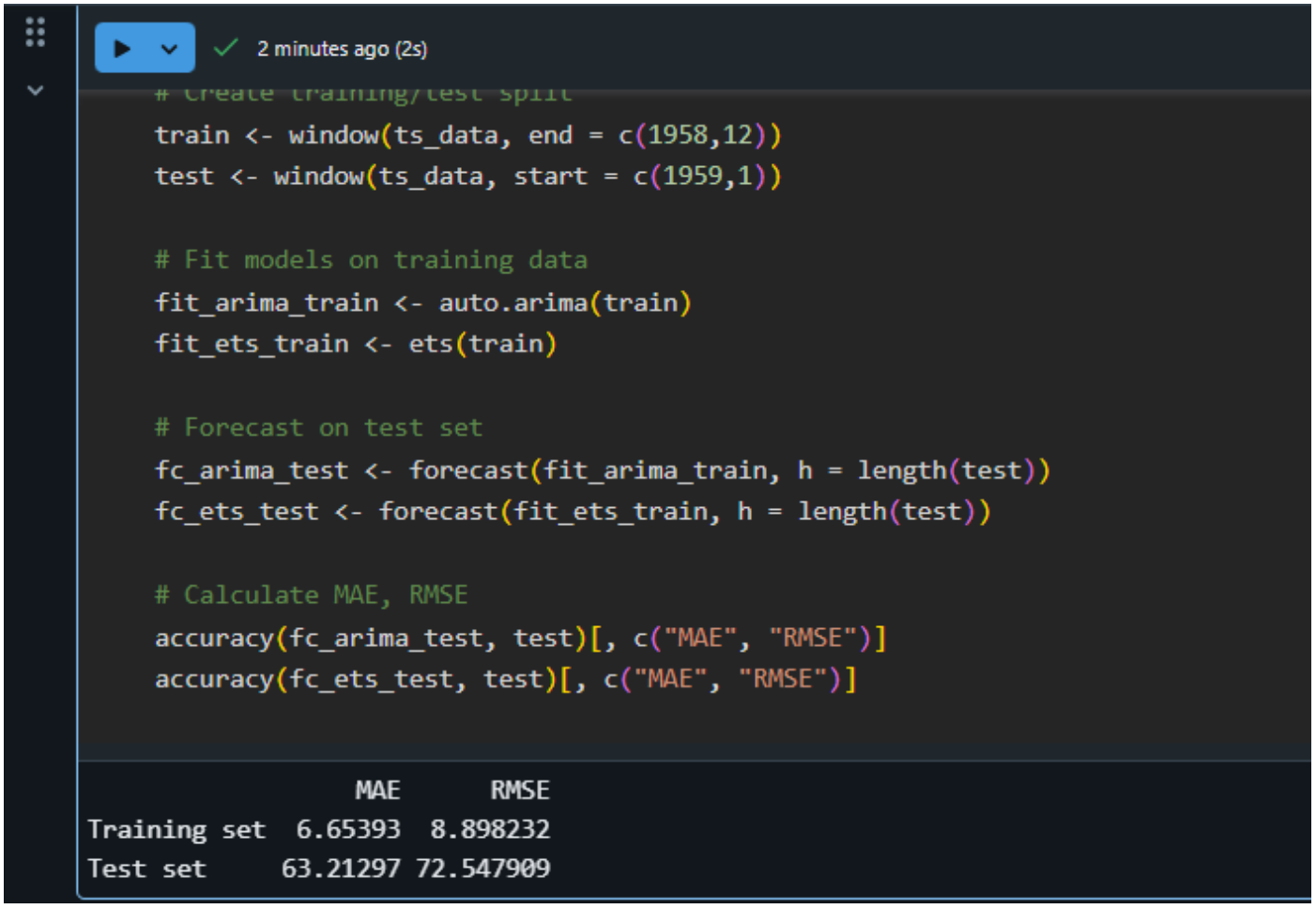



```
Warning in adf.test(ts_data) : p-value smaller than printed p-value
```

#Step 5 - **Fit ARIMA and ETS Models**

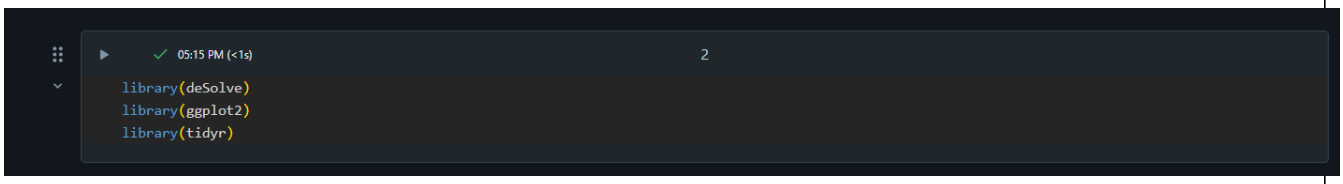


#Step 6 - Evaluate Model Performance

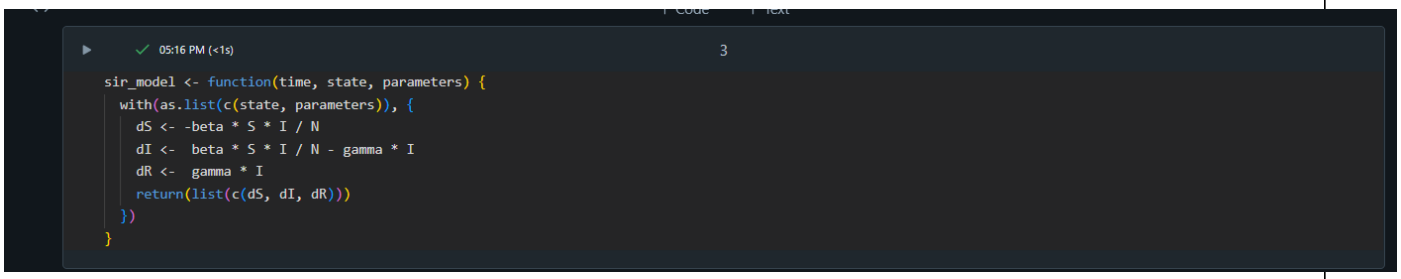


Task 2 – Disease Spread Simulation (SIR Model)

#Step 1 - Importing Libraries

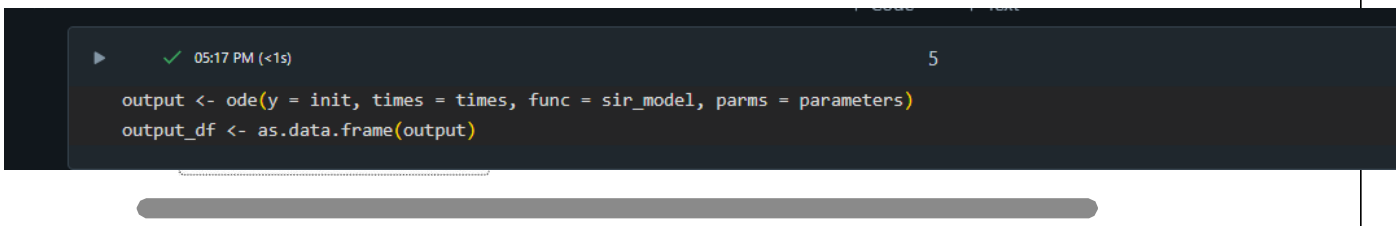


#Step 2 - Define the SIR
Differential Equations



#Step 3 - Set Initial Conditions and Parameters=

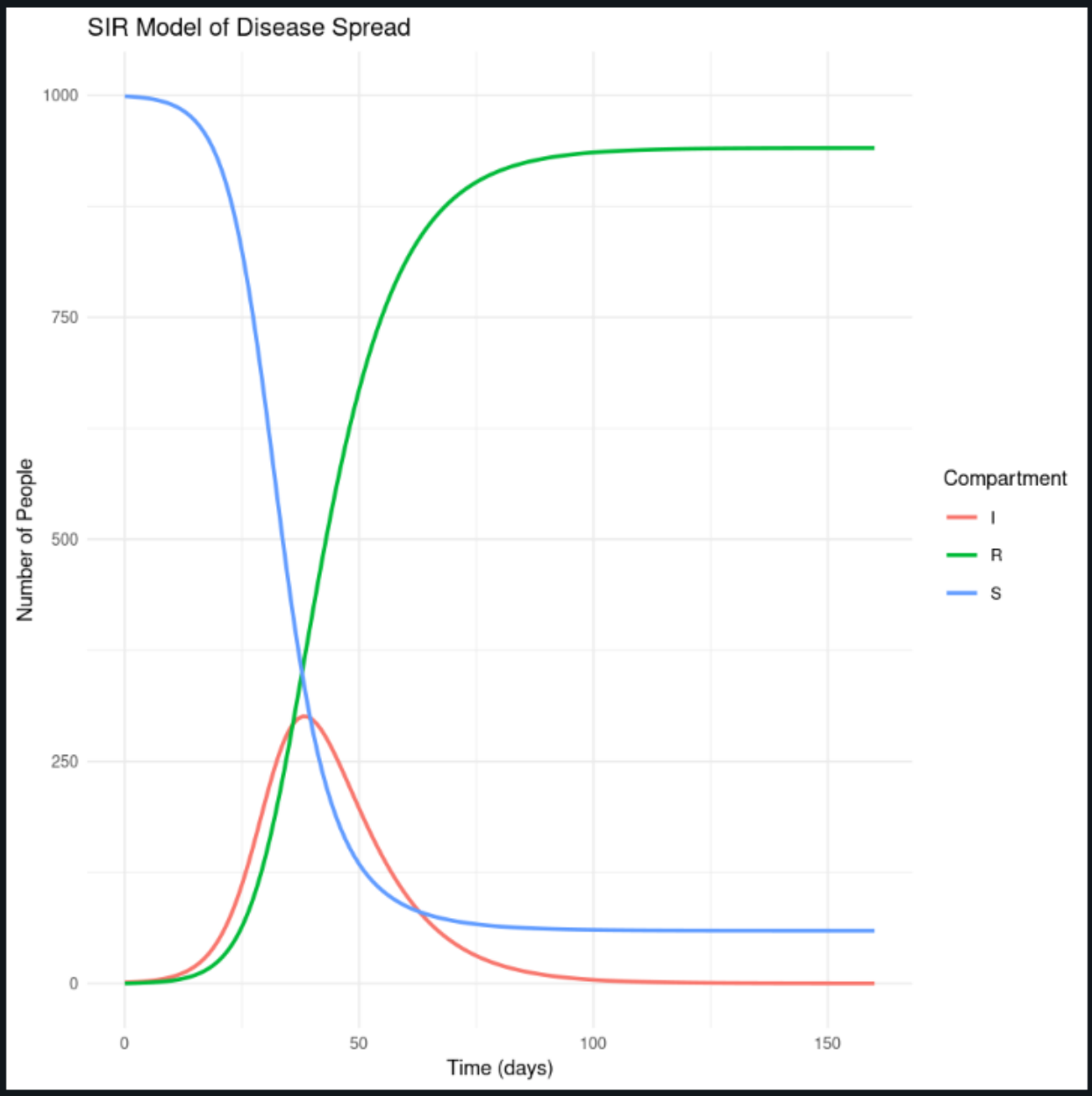
#Step 4 - Solve the Equations Using ode()



#Step 5- Evaluation

05:17 PM (1s) 6

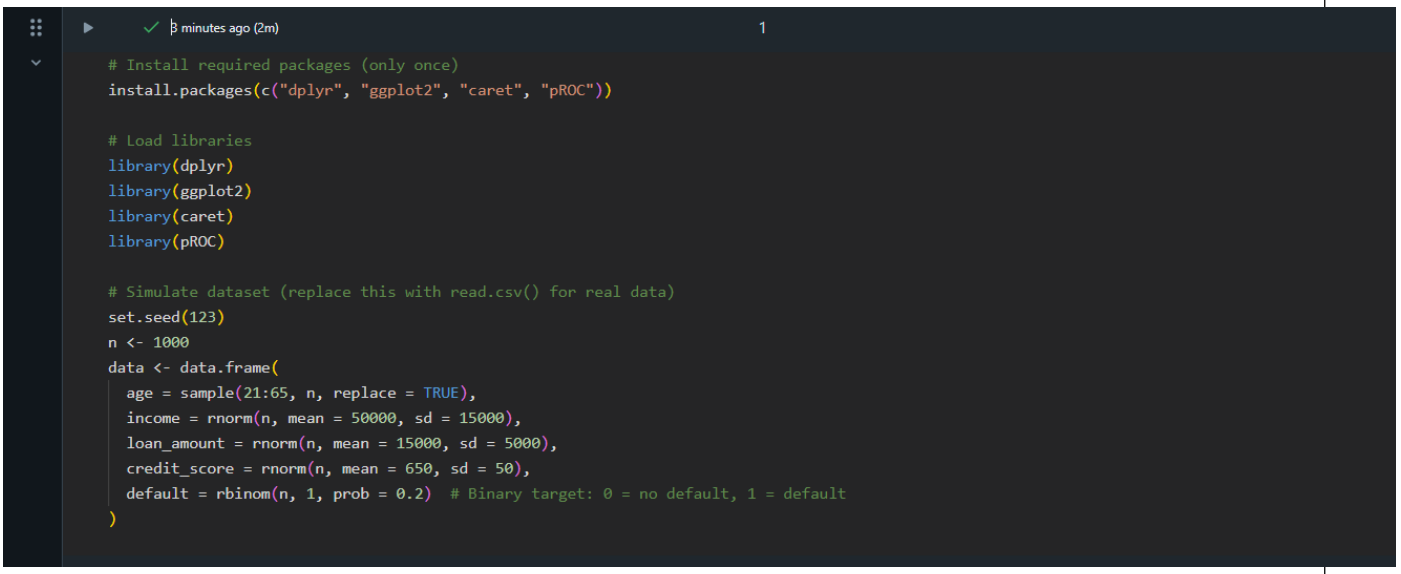
```
library(tidyr)
output_long <- pivot_longer(output_df, cols = c("S", "I", "R"), names_to = "Compartment", values_to = "Count")

ggplot(output_long, aes(x = time, y = Count, color = Compartment)) +
  geom_line(size = 1) +
  labs(title = "SIR Model of Disease Spread",
       x = "Time (days)", y = "Number of People") +
  theme_minimal()
```



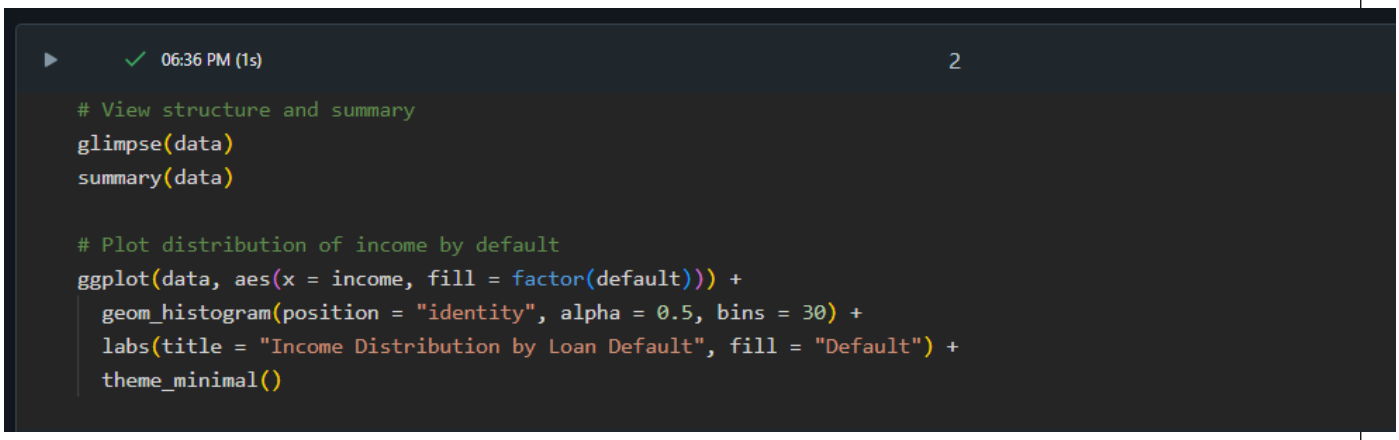

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.
```

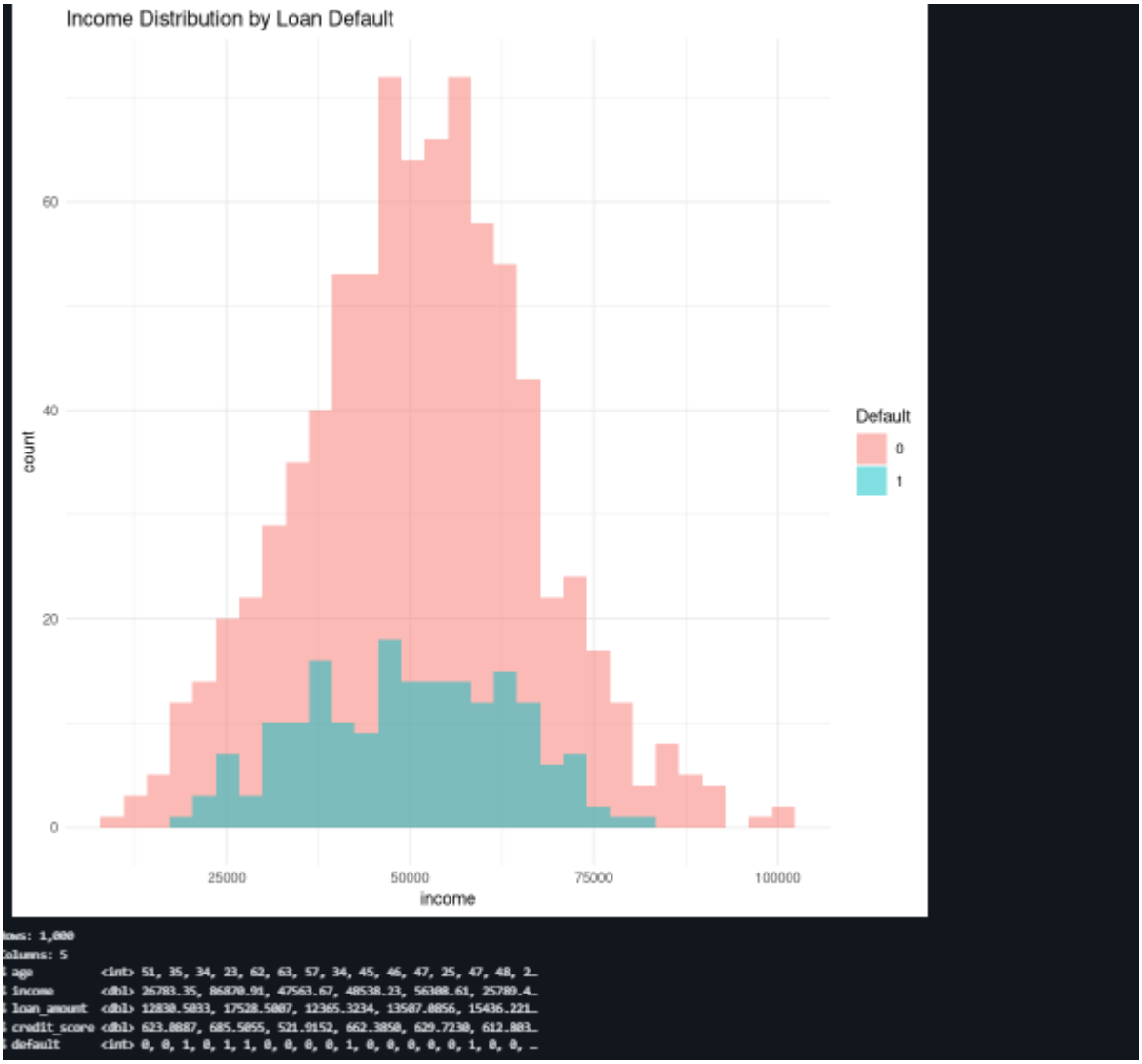
Task 3 – Credit Risk Modeling

#Step 1- Importing Libraries



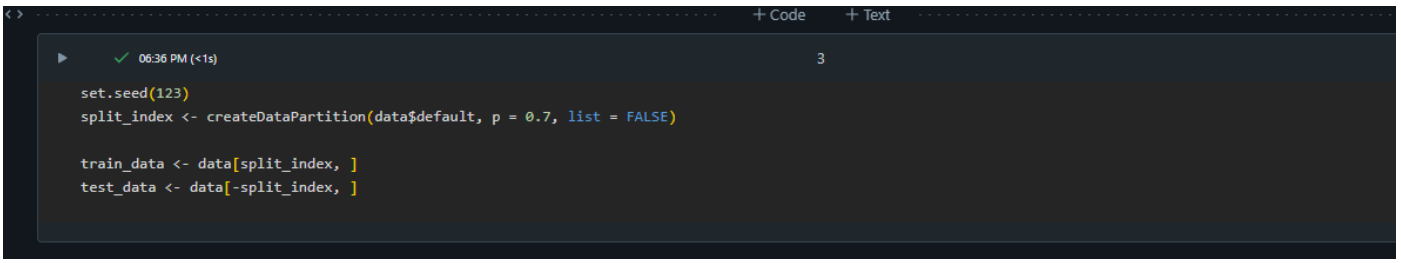
#Step 2- Load & Explore Dataset






```
ows: 1,000
olumns: 5
 age         <int> 51, 35, 34, 23, 62, 63, 57, 34, 45, 46, 47, 25, 47, 48, 2…
 income      <dbl> 26783.35, 86870.91, 47563.67, 48538.23, 56308.61, 25789.4…
 loan_amount <dbl> 12830.5033, 17528.5007, 12365.3234, 13507.0856, 15436.221…
 credit_score <dbl> 623.0887, 685.5055, 521.9152, 662.3850, 629.7230, 612.803…
 default     <int> 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, …
```

#Step 3 - Split Into Training and Test Sets



#Step 4 - Fit Logistic
Regression Model

```
06:36 PM (<1s)                                                    4

# Fit logistic regression model
logit_model <- glm(default ~ age + income + loan_amount + credit_score,
                   data = train_data, family = binomial)

# View model summary
summary(logit_model)
```

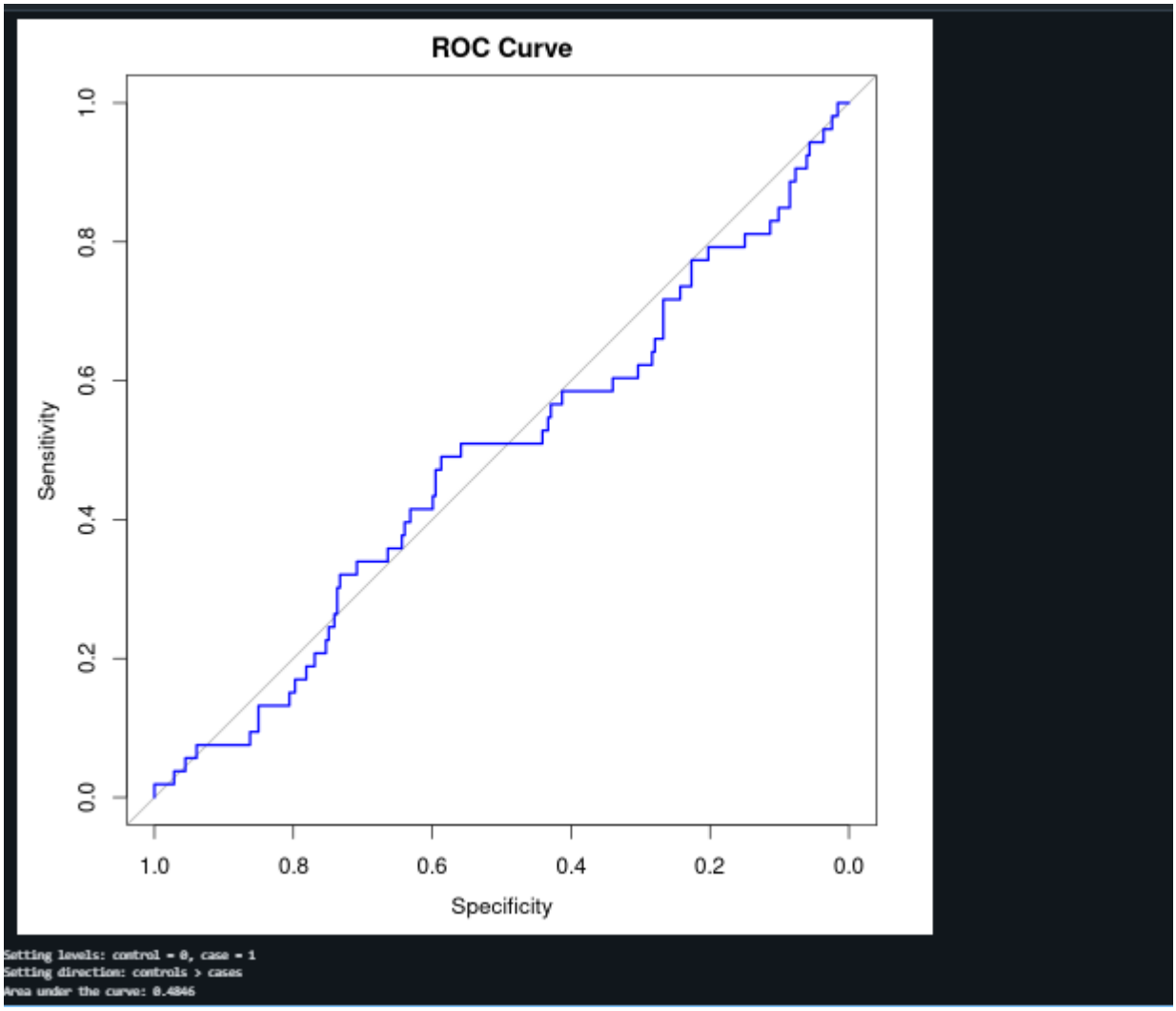
```
    family = binomial, data = train_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8036  -0.6697  -0.6242  -0.5681   1.9833

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.596e+00  1.396e+00  -1.144    0.253
age          -1.097e-02  7.749e-03  -1.415    0.157
income       -6.576e-06  6.213e-06  -1.058    0.290
loan_amount   3.315e-06  2.033e-05   0.163    0.870
credit_score  1.356e-03  1.972e-03   0.688    0.492

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 677.80  on 699  degrees of freedom
Residual deviance: 674.32  on 695  degrees of freedom
AIC: 684.32

Number of Fisher Scoring iterations: 4
```

#Step 5 - Predict Probabilities & Classify

```
06:36 PM (<1s)                                                    5

# Predict probabilities on test set
pred_probs <- predict(logit_model, test_data, type = "response")

# Set classification threshold
threshold <- 0.5
pred_class <- ifelse(pred_probs > threshold, 1, 0)
```

#Step 6 - Evaluate the Model

```
06:36 PM (<1s)                                                    7

# Plot ROC curve
roc_obj <- roc(test_data$default, pred_probs)
plot(roc_obj, main = "ROC Curve", col = "blue", lwd = 2)

# Calculate AUC
auc(roc_obj)
```


ROC Curve
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Setting levels: control = 0, case = 1
Setting direction: controls > cases
Area under the curve: 0.4846

Task 4 – Stock Price Analysis

#Step 1- Importing Libararies

```
1 minute ago (21s)                                                    1
install.packages(c("quantmod", "TTR"))  # Run only once
library(quantmod)
library(TTR)
```

#Step 2- **Fetch Stock Data from Yahoo Finance**

```
3 minutes ago (<1s)                                                   2
# Get Apple stock data from Yahoo Finance
getSymbols("AAPL", from = "2020-01-01", to = "2025-10-17", src = "yahoo")

# View first few rows
head(AAPL)
```

	AAPL.Open	AAPL.High	AAPL.Low	AAPL.Close	AAPL.Volume	AAPL.Adjusted
2020-01-02	74.0600	75.1500	73.7975	75.0875	135480400	72.71608
2020-01-03	74.2875	75.1450	74.1250	74.3575	146322800	72.00911
2020-01-06	73.4475	74.9900	73.1875	74.9500	118387200	72.58290
2020-01-07	74.9600	75.2250	74.3700	74.5975	108872000	72.24155
2020-01-08	74.2900	76.1100	74.2900	75.7975	132079200	73.40365
2020-01-09	76.8100	77.6075	76.5500	77.4075	170108400	74.96279

#Step 3- **Calculate Moving Averages (50-day and 200-day SMA)**

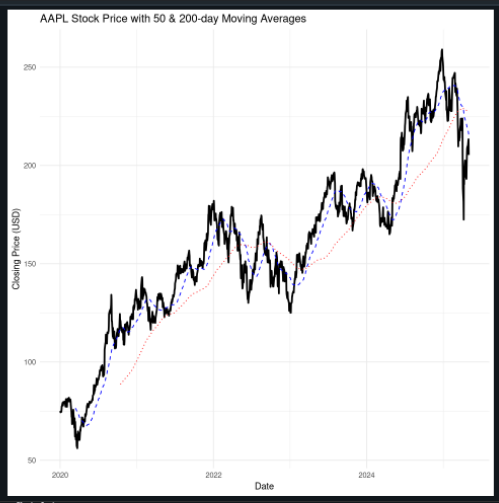
```
4 minutes ago (<1s)                                                   3
# Calculate 50-day and 200-day Simple Moving Averages
AAPL$SMA50 <- SMA(Cl(AAPL), n = 50)
AAPL$SMA200 <- SMA(Cl(AAPL), n = 200)
```

#Step 4 - Visualize Stock
Prices with Moving

```
08:50 PM (1s)                                                         4
# Basic chart with moving averages
chartSeries(AAPL,
            TA = "addSMA(n=50,col='blue'); addSMA(n=200,col='red')",
            theme = chartTheme("white"),
            name = "Apple Inc. (AAPL) with 50 & 200-day Moving Averages")
```

#Step 3- **Evaluation**

```
library(ggplot2)
library(dplyr)
library(tidyr)

# Convert to data frame
aapl_df <- data.frame(date = index(AAPL), coredata(AAPL))

# Plot
ggplot(aapl_df, aes(x = date)) +
  geom_line(aes(y = AAPL.Close), color = "black", size = 1) +
  geom_line(aes(y = SMA50), color = "blue", linetype = "dashed") +
  geom_line(aes(y = SMA200), color = "red", linetype = "dotted") +
  labs(title = "AAPL Stock Price with 50 & 200-day Moving Averages",
       x = "Date", y = "Closing Price (USD)") +
  theme_minimal()
```




```
first, last

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':
```

Task 5 -Sentiment Analysis on Tweets using R

```
# Install necessary packages (only run this once)
install.packages(c("tidytext", "dplyr", "ggplot2"))

# Load the libraries
library(tidytext)
library(dplyr)
library(ggplot2)

# Step 1: Sample Tweets (replace this with real tweets if needed)
tweets <- data.frame(
  text = c(
    "I love R programming!",
    "This project is difficult but rewarding.",
    "I hate bugs in the code!",
    "Learning data science is fun!",
    "Why is this not working?!"
  )
)

# Step 2: Text Cleaning and Sentiment Analysis
# Remove stop words and tokenize
data(stop_words)

words <- tweets %>%
  select(text) %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words)

# Use Bing lexicon for sentiment classification
bing <- get_sentiments("bing")

sentiment_data <- words %>%
  inner_join(bing, by = "word") %>%
  count(word, sentiment, sort = TRUE)

# Step 3: Summarize and Visualize Sentiment
sentiment_summary <- sentiment_data %>%
  group_by(sentiment) %>%
  summarise(total = sum(n))

# Plot the sentiment distribution
ggplot(sentiment_summary, aes(x = sentiment, y = total, fill = sentiment)) +
  geom_col() +
  labs(title = "Sentiment Analysis of Sample Tweets",
       x = "Sentiment",
       y = "Word Count") +
  theme_minimal()
```