Problem 2

Sharon Njeri

2024-03-05

```
# Load necessary libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.3.3

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(readxl)
```

Load the data

```
# Load the data
data <- read_excel("C:/Users/User/Downloads/Makert Prices 2022.xlsx")
library(readxl)
```

```
# merging entries with different case
data <- data %>%
  mutate(Seasons = case_when(
    Seasons == "dry" ~ "Dry",
    Seasons == "wet" ~ "Wet",
    TRUE ~ Seasons
  ))
```

```r
# Convert the Dates column to Date type
data$Dates <- as.Date(data$Dates, format="%m/%d/%Y")
```

```r
# Perform exploratory analysis
summary(data)
```

```
##      Dates              Seasons               Bull            Cow
##  Min.   :2009-08-01   Length:154         Min.   : 8500   Min.   :  966.7
##  1st Qu.:2012-10-08   Class :character   1st Qu.:45100   1st Qu.:21780.7
##  Median :2015-12-16   Mode  :character   Median :56504   Median :29150.4
##  Mean   :2015-12-17                      Mean   :53296   Mean   :28103.0
##  3rd Qu.:2019-02-22                      3rd Qu.:62418   3rd Qu.:33659.8
##  Max.   :2022-07-29                      Max.   :78000   Max.   :57613.0
##      Heifer          Steer
##  Min.   : 1667   Min.   :  766.7
##  1st Qu.:15083   1st Qu.:37421.7
##  Median :24462   Median :41150.2
##  Mean   :23831   Mean   :41550.1
##  3rd Qu.:29976   3rd Qu.:46531.3
##  Max.   :50000   Max.   :65411.5
```

```r
# Create dummy variables for seasons
data$Drought <- ifelse(data$Seasons == "Drought", 1, 0)
data$Wet <- ifelse(data$Seasons == "Wet", 1, 0)
data$Dry <- ifelse(data$Seasons == "Dry", 1, 0)
```

Modelling

```r
# creating a matrix of external regressors for the arima model
exreg= as.matrix(cbind(data$Drought,data$Wet,data$Dry))
colnames(exreg) <- c("Drought", "Wet", "Dry")
```

```r
# Fit ARIMA model with  season as external regressor.
Bull_model<- auto.arima(data$Bull, xreg = exreg )
Cow_model<- auto.arima(data$Cow, xreg = exreg )
Heifer_model<- auto.arima(data$Heifer, xreg = exreg )
Steer_model<- auto.arima(data$Steer, xreg = exreg )
```

```r
# Get the summary of each model
summary(Bull_model)
```

```
## Series: data$Bull
## Regression with ARIMA(2,1,3) errors
##
## Coefficients:
##          ar1     ar2      ma1      ma2      ma3    Drought       Wet        Dry
##       0.3166  0.1875  -0.2995  -0.3737  -0.1826  -1441.385  1300.872   997.2464
## s.e.  0.7497  0.5777   0.7402   0.5727   0.1640   5958.731  5161.074  5165.3424
##
```

```
## sigma^2 = 58716829:  log likelihood = -1581.74
## AIC=3181.47   AICc=3182.73   BIC=3208.74
##
## Training set error measures:
##                    ME    RMSE      MAE       MPE     MAPE      MASE
## Training set 875.0923 7435.41 5336.469 0.1838732 12.03243 0.9490133
##                      ACF1
## Training set -0.006111125
```

The model suggests that selling during a drought could lead to a decrease in price (coefficient: -1441.386), while selling during wet (coefficient: 1300.870) and dry (coefficient: 997.2451) seasons could lead to an increase in price. However, the standard errors are quite large, indicating a high degree of uncertainty in these estimates.

```
summary(Cow_model)
```

```
## Series: data$Cow
## Regression with ARIMA(1,0,1) errors
##
## Coefficients:
##          ar1     ma1  intercept    Drought       Wet        Dry
##       0.7934  0.5067   26520.57   -34.9539  1309.246    94.7516
## s.e.  0.0564  0.0798    3692.29  3359.5915  2895.063  2887.6079
##
## sigma^2 = 20102868:  log likelihood = -1511.3
## AIC=3036.6   AICc=3037.37   BIC=3057.86
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE    MAPE      MASE        ACF1
## Training set 150.0031 4395.411 2803.349 -9.204389 18.4434 0.8836972 -0.02841004
```

The model suggests that selling during a drought could lead to a slight decrease in price (coefficient: -34.9539), while selling during wet (coefficient: 1309.246) and dry (coefficient: 94.7516) seasons could lead to an increase in price. The standard errors are also quite large for this model.

```
summary(Heifer_model)
```

```
## Series: data$Heifer
## Regression with ARIMA(4,0,0) errors
##
## Coefficients:
##          ar1      ar2     ar3      ar4  intercept   Drought       Wet       Dry
##       1.3521  -0.6355  0.4462  -0.2325  18235.242  3338.680  3454.366  3531.346
## s.e.  0.0877   0.1532  0.1561   0.0922   4329.108  2497.617  2183.864  2162.105
##
## sigma^2 = 11563277:  log likelihood = -1468.12
## AIC=2954.24   AICc=2955.49   BIC=2981.58
##
## Training set error measures:
##                    ME    RMSE      MAE       MPE    MAPE      MASE        ACF1
## Training set 130.3219 3310.98 2203.358 -3.959579 14.0376 0.8105078 -0.09775363
```

Cow: The model suggests that selling during a drought could lead to a slight decrease in price (coefficient: -34.9539), while selling during wet (coefficient: 1309.246) and dry (coefficient: 94.7516) seasons could lead to an increase in price. The standard errors are also quite large for this model.

```
summary(Steer_model)
```

```
## Series: data$Steer
## Regression with ARIMA(4,1,0) errors
##
## Coefficients:
##          ar1      ar2      ar3     ar4    Drought        Wet        Dry
##       0.0962  -0.1479  -0.3006  0.1206  -5232.987  -3645.549  -3984.431
## s.e.  0.0829   0.0821   0.0816  0.0857   4483.666   3953.997   3993.225
##
## sigma^2 = 36672653:  log likelihood = -1546.17
## AIC=3108.34   AICc=3109.34   BIC=3132.58
##
## Training set error measures:
##                    ME     RMSE     MAE       MPE     MAPE      MASE        ACF1
## Training set 332.7419 5896.404 3933.55 0.3555691 10.47228 0.9506942 0.001327854
```

The model suggests that selling during all seasons could lead to a decrease in price (coefficients: -5232.987, -3645.549, and -3984.431 for drought, wet, and dry seasons respectively). The standard errors are quite large, indicating a high degree of uncertainty.

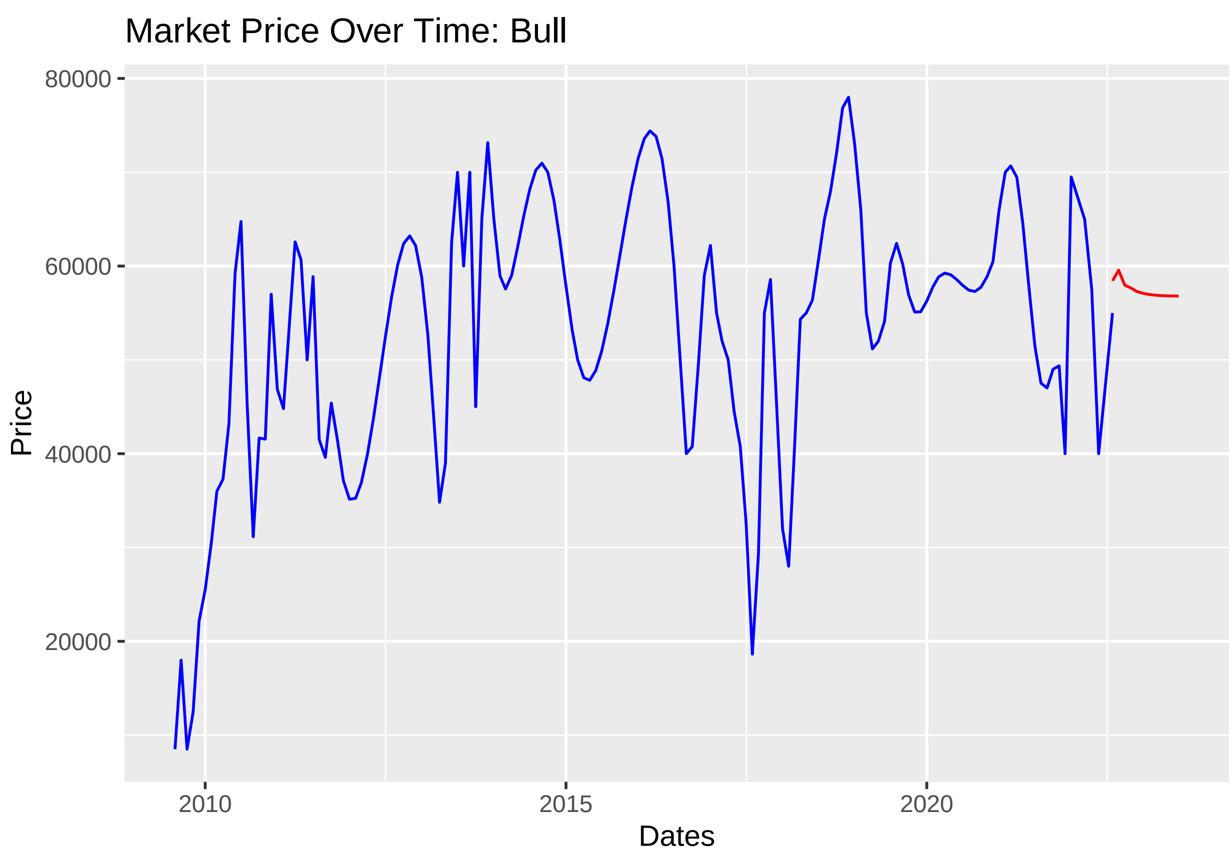
```
# Get the seasons for the last 12 months
last_12_seasons <- tail(exreg, 12)

# Create a future exreg matrix based on these seasons
future_exreg <- model.matrix(~last_12_seasons-1)
colnames(future_exreg) <- c("Drought", "Wet", "Dry")

# Forecast the next 12 months
Bull_forecast <- forecast(Bull_model, h=12, xreg=future_exreg)
Cow_forecast <- forecast(Cow_model, h=12, xreg=future_exreg)
Heifer_forecast <- forecast(Heifer_model, h=12, xreg=future_exreg)
Steer_forecast <- forecast(Steer_model, h=12, xreg=future_exreg)
```

Plotting the forecasts.

1. Bull

```
# Create a data frame for the historical data
historical_data <- data.frame(
  Dates = data$Dates,
  Price = data$Bull
)

# Create a data frame for the forecasted data
forecast_data <- data.frame(
  Dates = seq(max(data$Dates), by = "month", length.out = 12),
```

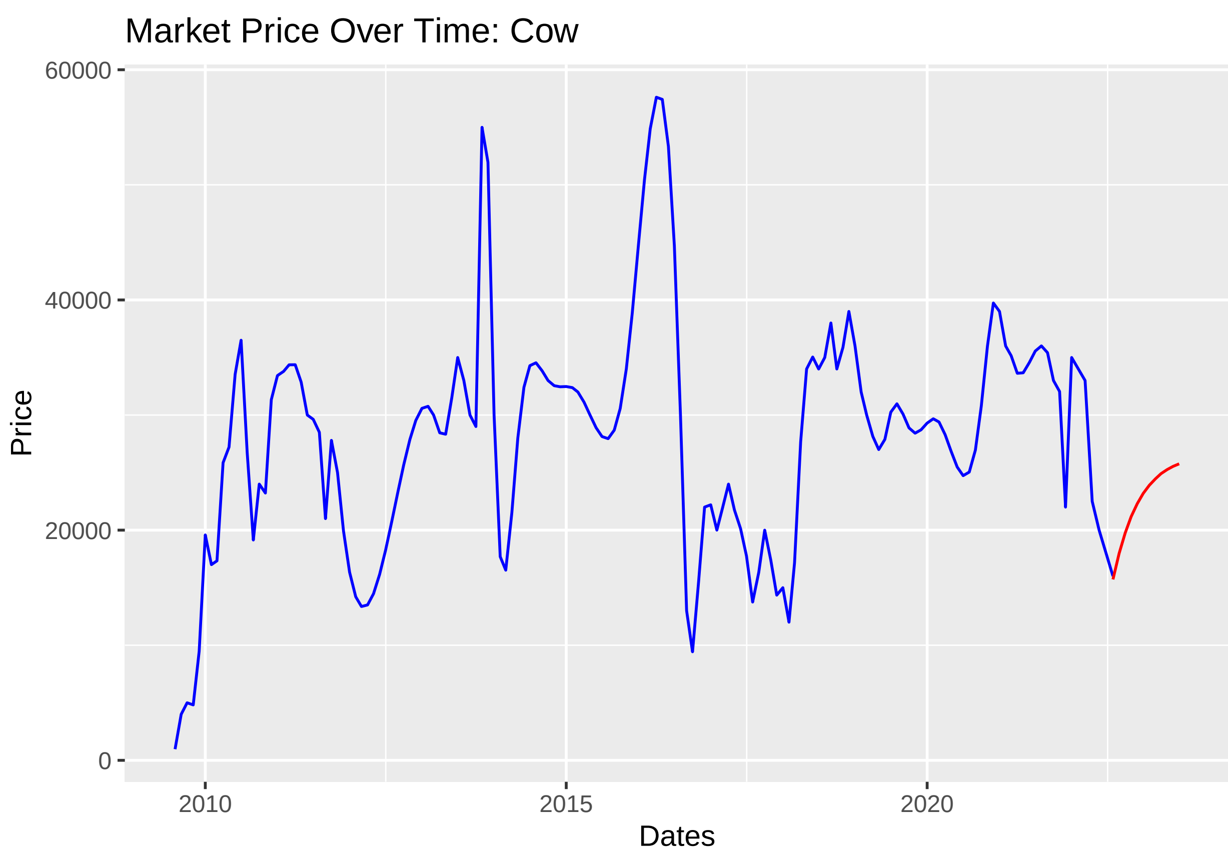
```r
  Price = as.numeric(Bull_forecast$mean)
)

# Combine the historical and forecasted data
combined_data <- rbind(historical_data, forecast_data)

# Convert 'Dates' to Date class
combined_data$Dates <- as.Date(combined_data$Dates)

ggplot() +
  geom_line(data = historical_data, aes(x=Dates, y=Price), color = "blue") +
  geom_line(data = forecast_data, aes(x=Dates, y=Price), color = "red") +
  labs(x = "Dates", y = "Price", title = "Market Price Over Time: Bull")
```



2. Cow

```r
# Create a data frame for the historical data
historical_data_cow <- data.frame(
  Dates = data$Dates,
  Price = data$Cow
)

# Create a data frame for the forecasted data
forecast_data_cow <- data.frame(
  Dates = seq(max(data$Dates), by = "month", length.out = 12),
```

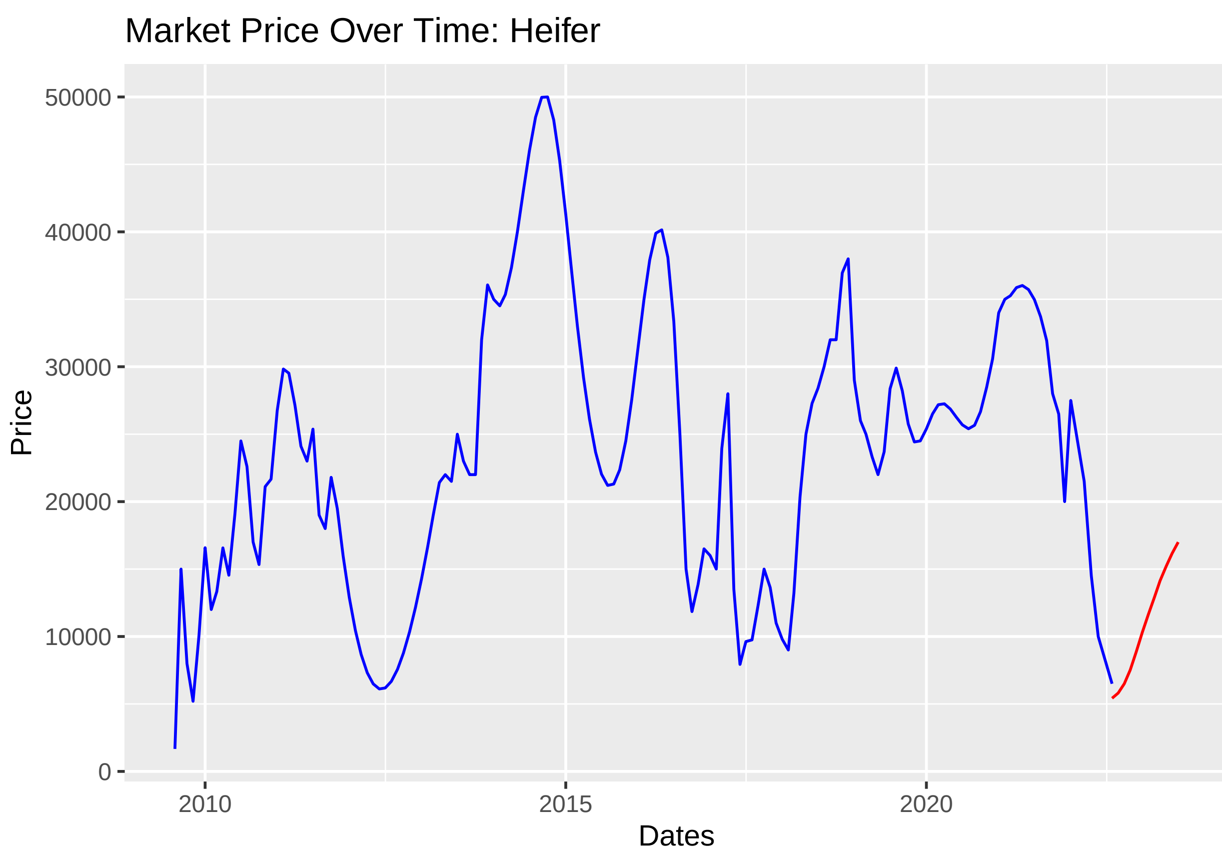
```
  Price = as.numeric(Cow_forecast$mean)
)

# Combine the historical and forecasted data
combined_data_cow <- rbind(historical_data_cow, forecast_data_cow)

# Convert 'Dates' to Date class
combined_data_cow$Dates <- as.Date(combined_data_cow$Dates)

ggplot() +
  geom_line(data = historical_data_cow, aes(x=Dates, y=Price), color = "blue") +
  geom_line(data = forecast_data_cow, aes(x=Dates, y=Price), color = "red") +
  labs(x = "Dates", y = "Price", title = "Market Price Over Time: Cow")
```



3. Heifer

```
# Create a data frame for the historical data
historical_data_heifer <- data.frame(
  Dates = data$Dates,
  Price = data$Heifer
)

# Create a data frame for the forecasted data
forecast_data_heifer <- data.frame(
  Dates = seq(max(data$Dates), by = "month", length.out = 12),
```

```r
  Price = as.numeric(Heifer_forecast$mean)
)

# Combine the historical and forecasted data
combined_data_heifer <- rbind(historical_data_heifer, forecast_data_heifer)

# Convert 'Dates' to Date class
combined_data_heifer$Dates <- as.Date(combined_data_heifer$Dates)

ggplot() +
  geom_line(data = historical_data_heifer, aes(x=Dates, y=Price), color = "blue") +
  geom_line(data = forecast_data_heifer, aes(x=Dates, y=Price), color = "red") +
  labs(x = "Dates", y = "Price", title = "Market Price Over Time: Heifer")
```



4. Steer

```r
# Steer
# Create a data frame for the historical data
historical_data_steer <- data.frame(
  Dates = data$Dates,
  Price = data$Steer
)

# Create a data frame for the forecasted data
forecast_data_steer <- data.frame(
  Dates = seq(max(data$Dates), by = "month", length.out = 12),
```

```
  Price = as.numeric(Steer_forecast$mean)
)

# Combine the historical and forecasted data
combined_data_steer <- rbind(historical_data_steer, forecast_data_steer)

# Convert 'Dates' to Date class
combined_data_steer$Dates <- as.Date(combined_data_steer$Dates)

ggplot() +
  geom_line(data = historical_data_steer, aes(x=Dates, y=Price), color = "blue") +
  geom_line(data = forecast_data_steer, aes(x=Dates, y=Price), color = "red") +
  labs(x = "Dates", y = "Price", title = "Market Price Over Time: Steer")
```