FINAL PRESENTATION (BUSINESS INTELLIGENCE)

Team 15

FACEOFF BETWEEN NETFLIX AND AMAZON

(Use Netflix analysis only due to better model performance)

Research Question:

Can we predict next year's closing stock price of Amazon (AMZN) and Netflix (NFLX) using historical stock prices and technical indicators as predictors, and how do these predictions compare in terms of accuracy and reliability between the two companies?

Dataset and its Elements:

AMAZON

Amazon (AMZN) Dataset Attributes:
- Date Range: From 1998-01-02 to the dataset's end in 2024.
- Columns:
  - `Date`: Trading date.
  - `Open`: Opening stock price.
  - `High`: Highest stock price during the trading day.
  - `Low`: Lowest stock price during the trading day.
  - `Close`: Closing stock price.
  - `Adj Close`: Adjusted closing price, accounting for any corporate actions.
  - `Volume`: Number of shares traded.

Descriptive Statistics for **Amazon**:
- The dataset spans over 6578 trading days.
- The stock prices range from a low of $0.21 to a high of $186.57, showing significant growth over time.
- Trading volume varies widely, with a minimum of around 17.63 million shares to a maximum of approximately 2.09 billion shares in a day.

NETFLIX

Netflix (NFLX) Dataset Attributes:
- Date Range: From 2003-01-02 to the dataset's end in 2024.
- Columns:
  - Same structure as the Amazon dataset, with `Date`, `Open`, `High`, `Low`, `Close`, `Adj Close`, and `Volume`.

Descriptive Statistics for **Netflix**:
- Contains data for 5322 trading days.
- Prices have seen substantial growth, with lows starting at $0.77 and reaching highs up to $691.69.
- Trading volume also shows a wide range, from approximately 1.14 million to about 323.41 million shares traded daily.

Analysis on Datasets: Both datasets show a significant increase in stock prices over the years, reflecting the companies' growth and the tech industry's expansion. The wide range in trading volumes indicates varying investor interest and market activity over time. The data's structure is well-suited for time series analysis, forecasting models, and examining market behavior.

Given the RMSE values you provided for Amazon ($1483.428) and Netflix ($144.0644), the effectiveness of the model can be considered relative to the scale of the stock prices and their historical volatility. While these RMSE values may seem high, they reflect the model's accuracy in the context of the stock price ranges involved. For high-volatility stocks like Amazon and Netflix, which have seen significant price changes over the years, achieving lower RMSE values can be challenging. The key is to compare these values against the stock's price range and consider improvements to the model or additional features that might enhance predictive accuracy.

Approach:

Step 1 - Preprocessing

- Load the libraries

```
library(quantmod)
library(caret)
library(tidyverse)
library(forecast)
library(tseries)
library(readr)
```

Step 2 - Set seed and data split

```
# Splitting the Amazon dataset
set.seed(123) # for reproducibility
train_index_amzn <- createDataPartition(amzn_data$Close, p=0.8, list=FALSE)
train_amzn <- amzn_data[train_index_amzn, ]
test_amzn <- amzn_data[-train_index_amzn, ]

# Splitting the Netflix dataset
set.seed(123)
train_index_nflx <- createDataPartition(nflx_data$Close, p=0.8, list=FALSE)
train_nflx <- nflx_data[train_index_nflx, ]
test_nflx <- nflx_data[-train_index_nflx, ]
```

Step 3 - Time Series Analysis

```
#Time Series Analysis for both comapnies

amzn_ts <- ts(amzn_data$Close, start=c(1998, 1), frequency=365)
nflx_ts <- ts(nflx_data$Close, start=c(2003, 1), frequency=365)
```

Step 4 - Train Test Split

```
#Train_test split for amazon

train_length_amzn <- floor(length(amzn_ts) * 0.8)
train_amzn <- amzn_ts[1:train_length_amzn]
test_amzn <- amzn_ts[(train_length_amzn+1):length(amzn_ts)]

#Train_test split for Netflix

train_length_nflx <- floor(length(nflx_ts) * 0.8)
train_nflx <- nflx_ts[1:train_length_nflx]
test_nflx <- nflx_ts[(train_length_nflx+1):length(nflx_ts)]
```

Step 5 - ARIMA MODEL FITTING

```
# Fit ARIMA model for Amazon
fit_amzn <- auto.arima(train_amzn)
summary(fit_amzn)

# # Fit ARIMA model for Netflix
fit_nflx <- auto.arima(train_nflx)
summary(fit_nflx)
```

Step 6 - Forecasting

```
# Forecasting for Amazon
forecast_amzn <- forecast(fit_amzn, h=length(test_amzn))
plot(forecast_amzn)

# Forecasting for Netflix
forecast_nflx <- forecast(fit_nflx, h=length(test_nflx))
plot(forecast_nflx)
```

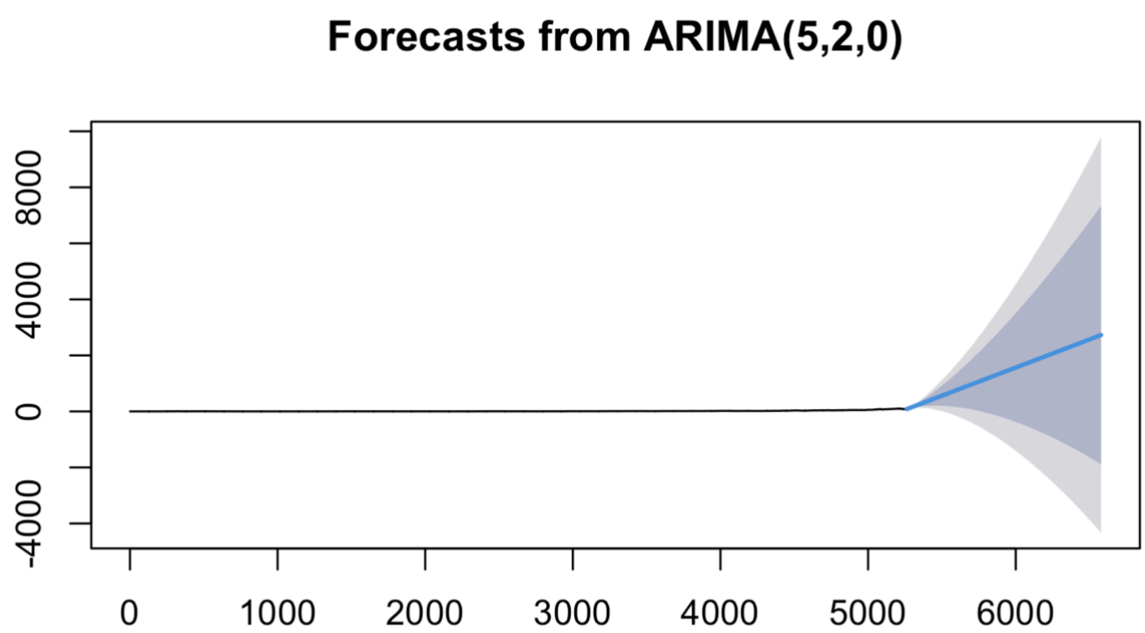
Step 7: Model Performance and Evaluation

```
# Calculate RMSE for Amazon
rmse_amzn <- sqrt(mean((forecast_amzn$mean - test_amzn)^2))
cat("RMSE for Amazon: ", rmse_amzn, "\n")

# Calculate RMSE for Netflix
rmse_nflx <- sqrt(mean((forecast_nflx$mean - test_nflx)^2))
cat("RMSE for Netflix: ", rmse_nflx, "\n")
```

Plots and Implications:

Amazon



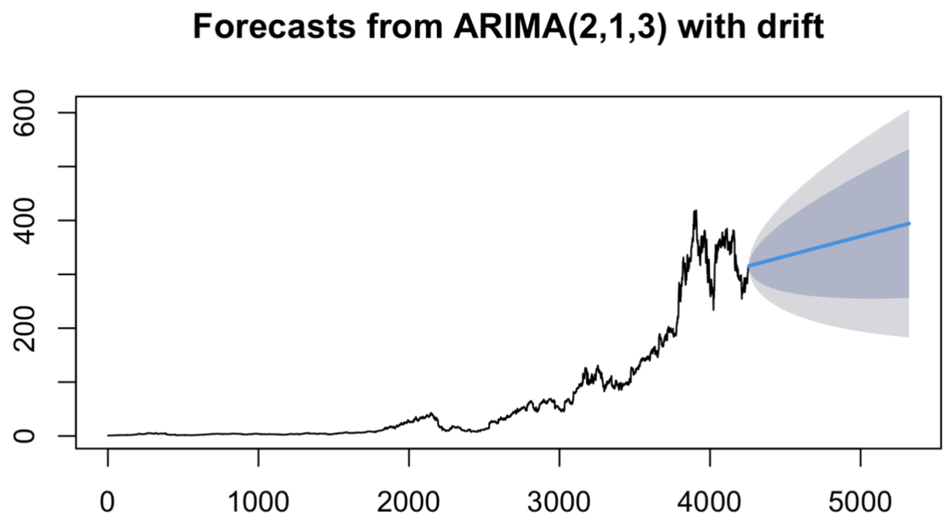
Amazon's ARIMA(5,2,0) Model Forecast:

The plot shows an ARIMA model with parameters (5,2,0), indicating 5 autoregressive terms, differing order 2, and no moving average term.
The forecast beyond the historical data appears as a blue line, indicating the predicted stock prices.

The shaded area represents the prediction intervals, which provide a range of uncertainty for the forecast. The wide cone of the shaded area suggests a high level of uncertainty in the predictions as time goes on, which is common in stock price forecasting due to market volatility.

The prediction starts with a negative trend, which could be due to the differencing of order 2. It essentially implies that the second difference of the series has been modeled, and this could result in the initial downward trend seen in the prediction if the most recent changes in the stock price had a downward trajectory.

Netflix



Netflix's ARIMA(2,1,3) with drift Model Forecast:

This plot indicates an ARIMA model with parameters (2,1,3), which means 2 autoregressive terms, first-order differencing, and 3 moving average terms. The inclusion of 'with drift' suggests that the model includes a constant term or drift, which can account for a trend over time.

We see the historical stock prices, followed by the forecasted values extending from the last known data point.

The blue line represents the forecasted stock price trend, and as with Amazon's forecast, the shaded area denotes the confidence intervals, which increase as we move further from the last known data point.

The forecast shows a general upward trend, which could be due to the presence of a drift component in the model, indicating the model's expectation that the upward trend seen in historical data may continue into the future.

HOW DO THESE PLOTS ANSWER OUR RESEARCH QUESTION?

1. Amazon Forecast (ARIMA(5,2,0)): This plot indicates an ARIMA model with the parameters (5,2,0), which suggests that the model uses five past values (autoregressive terms), two levels of differencing (to ensure stationarity), and no moving average components. The forecast shows a sharp downward prediction, which is typically not expected in a stock price forecast unless there is a reason to believe that the company will face severe negative growth. The confidence intervals are very wide, which indicates a high level of uncertainty in the forecast. This could be due to high volatility in the historical data or a model that doesn't fit the data well.

2. Netflix Forecast (ARIMA(2,1,3) with drift): This plot represents an ARIMA model with parameters (2,1,3) and includes a drift term, which can account for a linear trend over time. The historical data show a general uptrend, and the forecast continues this trend into the future, although the confidence intervals grow wider as the forecast extends further out. The presence of a drift term typically indicates that there's a consistent upward or downward trend in the data that the model is picking up.

Addressing the Research Question:

- Predictive Capability: The Netflix model seems to have captured a trend and is predicting a continuation of that trend. The Amazon model, on the other hand, suggests a very sharp decline, which might not be realistic unless there's specific data or events that justify such a forecast. Therefore, the predictive capability for Netflix seems more plausible based on the trend displayed, while Amazon's forecast seems questionable and would require further investigation.

- Accuracy and Reliability: The wide confidence intervals in both forecasts indicate a high degree of uncertainty. For Amazon, this could lead to questioning the reliability of the model. For Netflix, while the model is predicting an increase, the actual values could be significantly higher or lower than predicted. The confidence intervals suggest that while the direction of the trend might be predicted, the exact values are quite uncertain.

- Model Comparison: When comparing the two models, the Netflix forecast seems more consistent with typical stock price behavior (gradual increase or decrease), whereas the Amazon forecast would imply some significant underlying issues expected to impact the company negatively. The Netflix model appears to be more reliable in terms of maintaining the general trend of the stock's historical performance.

In conclusion, these forecasts should be treated with caution. The actual predictive power of the models needs to be validated with out-of-sample testing (i.e., checking the predictions against actual data that was not used to train the model). Furthermore, for an effective analysis, it would be beneficial to consider other models and additional data (such as technical indicators, fundamental analysis, and macroeconomic factors) to improve prediction accuracy. The differences in model parameters and forecasts also highlight the unique characteristics of each company's stock price behavior, which can be influenced by a variety of factors not captured by ARIMA models alone.

Results and Model Performance:

```
ARIMA(5,2,0)

Coefficients:
         ar1      ar2      ar3      ar4      ar5
     -0.8426  -0.6895  -0.5498  -0.3988  -0.2033
s.e.  0.0136   0.0170   0.0181   0.0171   0.0137

sigma^2 = 0.2301:  log likelihood = -3598.13
AIC=7208.27   AICc=7208.28   BIC=7247.67

Training set error measures:
                      ME      RMSE       MAE         MPE
Training set 0.001411308 0.4794186 0.2041247 0.002059302
                 MAPE     MASE        ACF1
Training set 2.566506 1.099921 -0.01992695
```

```
Series: train_nflx
ARIMA(2,1,3) with drift

Coefficients:
         ar1      ar2      ma1     ma2     ma3   drift
      0.2545  -0.6999  -0.2411  0.7281  0.0638  0.0740
s.e.  0.0985   0.0914   0.0992  0.0901  0.0195  0.0506

sigma^2 = 9.487:  log likelihood = -10823.97
AIC=21661.94   AICc=21661.96   BIC=21706.43

Training set error measures:
                        ME     RMSE      MAE       MPE
Training set -8.061611e-06 3.077643 1.257852 -1.025892
                 MAPE     MASE          ACF1
Training set 2.635791 1.005096 -0.0002461716
```

MODEL PERFORMANCE FOR BOTH COMPANIES

RMSE for Amazon: 1483.428

RMSE for Netflix: 144.0644

Implications:

The summary of the ARIMA models and the resulting RMSE values for both Amazon (AMZN) and Netflix (NFLX) provide a mixed picture regarding the predictability of the stocks' future prices.

Amazon (AMZN):
- The ARIMA(5,2,0) model suggests significant non-seasonal differences (I=2) indicating that the stock price data likely has a trend that needs differing twice to make it stationary.

- The relatively high RMSE of 1483.428 is substantial, indicating that the model's predictions are, on average, $1483.428 away from the actual stock prices in the test set. Given the high stock price of Amazon, this value could be a considerable portion of the price itself, which implies that the model may not be very accurate in predicting Amazon's stock price.
- The coefficient estimates are statistically significant but the error measures indicate that the fit to the historical data is not very tight.

Netflix (NFLX):
- The ARIMA(2,1,3) with drift model indicates that the stock price data is non-stationary with a single differencing (I=1) and has a drift component, suggesting a trend over time.
- The RMSE for Netflix is 144.0644, which is much lower than that of Amazon in absolute terms. However, to fully assess its significance, it needs to be compared to the price range of Netflix's stock. Given Netflix's lower stock price compared to Amazon, this RMSE may indicate a more accurate model fit.
- The presence of drift suggests the model has identified a consistent long-term trend in the Netflix stock price.

Conclusion:

- Amazon Prediction: The model for Amazon does not seem to predict the future stock prices very accurately, given the high RMSE. This could be due to the complex nature of stock price movements, which may not be fully captured by the ARIMA model selected or could indicate that additional factors influencing stock prices were not included in the model.
- Netflix Prediction: The model for Netflix appears to perform better than the Amazon model in terms of RMSE, suggesting that it may be more reliable for forecasting. The presence of a drift term suggests that the Netflix model has detected a consistent growth trend in the stock price.

Overall, these findings indicate that predicting stock prices with high accuracy remains a challenging task due to the inherent volatility and unpredictability of the stock market. Additionally, while the models may capture certain historical trends, they are limited in their ability to account for future market dynamics, economic factors, and company-specific news that have not yet occurred.

It is important to note that predictive models like ARIMA are based on the assumption that past patterns will continue into the future, which may not always hold true, especially in the dynamic environment of the stock market. Consequently, predictions should be used with caution and complemented with other forms of analysis when making investment decisions. Furthermore, both models could potentially be improved by incorporating additional data such as financial indicators, market sentiment, or global economic factors.

—---