I. Introduction

* Introduce the problem of predicting future stock prices using time series market data.
* Formulate a hypothesis about which predictive models are most effective.

II. Data Collection and Preparation

* Discuss the data sources used to test the hypothesis.
* Describe the process used to collect and clean the data, including any feature engineering techniques.

III. Analysis

* Compare the effectiveness of different predictive models, such as ARIMA, LSTM, and Prophet.
* Demonstrate a deep understanding of these models and explain why they were chosen.
* Discuss the evaluation metrics used to assess model performance.

IV. Scientific Method and Documentation

* Explain the scientific method used to compare the different predictive models.
* Discuss the importance of documenting the process and results, including any limitations or assumptions.
* Highlight attention to detail and communication skills in the documentation.

V. Findings and Conclusion

* Summarize the insights and conclusions drawn from the analysis.
* Discuss which predictive models were most effective and why.
* Demonstrate creativity and communication skills by presenting the findings in an engaging way.

VI. Discussion and Questions

* Be prepared to answer questions about the process, results, and findings, including any limitations or potential applications of the research.

I. Introduction

* Explain the goal of the project: to investigate different predictive models using the same time series market data.
* Introduce the problem of predicting future stock prices using time series market data.
* Formulate a hypothesis about which predictive models are most effective.

II. Data Collection and Preparation

* Discuss the data sources used to test the hypothesis.
* Describe the process used to collect and clean the data, including any feature engineering techniques.
* Collect historical data of the stock prices over a period of time (e.g. daily, weekly, monthly).
* Check for missing values and outliers and handle them appropriately (e.g. impute, drop, replace).
* Transform the data into a stationary series if needed (e.g. differencing, log transformation).
* Discuss the need for handling missing values and outliers appropriately.
* Explain the need for transforming data into a stationary series if needed.
* Split the data into train and test sets.

III. Exploratory Data Analysis

* Plot the data and observe the patterns and trends (e.g. seasonality, cyclicity, trend).
* Calculate descriptive statistics and summary measures (e.g. mean, median, standard deviation, skewness, kurtosis).
* Perform correlation analysis and test for autocorrelation and partial autocorrelation (e.g. ACF, PACF plots).
* Discuss the need for performing correlation analysis and testing for autocorrelation and partial autocorrelation.

IV. Analysis

* Model testing
  + Discuss the different models that can be used, such as ARIMA, SARIMA, ETS, and LSTM.
  + Explain the process of specifying the model parameters or hyperparameters and fitting the model on the train set.
  + Discuss the importance of evaluating the model performance on the test set using appropriate metrics.
* Model selection
* Explain the process of comparing different models or variations of models based on their performance metrics and choosing the best one.
* Discuss the need for checking for model assumptions and diagnostics, such as residuals analysis, to validate the model.
* Model evaluation and prediction
* Discuss the importance of applying the selected model to forecast future values of the stock prices using confidence intervals or prediction intervals.
* Explain the need for updating or retraining the model periodically as new data becomes available.
* Compare the effectiveness of different predictive models, such as ARIMA, LSTM, and Prophet.
* Demonstrate a deep understanding of these models and explain why they were chosen.
* Discuss the evaluation metrics used to assess model performance.
* Choose a suitable model or models based on the characteristics of the data (e.g. ARIMA, SARIMA, ETS, LSTM).
* Specify the model parameters or hyperparameters and fit the model on the train set.
* Evaluate the model performance on the test set using appropriate metrics (e.g. MAE, MSE, RMSE).
* Compare different models or variations of models based on their performance metrics and choose the best one.
* Check for model assumptions and diagnostics (e.g. residuals analysis) and validate the model.

VII. Findings and Conclusion

* Summarize the insights and conclusions drawn from the analysis.
* Discuss which predictive models were most effective and why.
* Demonstrate creativity and communication skills by presenting the findings in an engaging way.

V. Scientific Method and Documentation

* Explain the scientific method used to compare the different predictive models.
* Discuss the importance of documenting the process and results, including any limitations or assumptions.

VI. Code

* Apply the selected model to forecast future values of the stock prices using confidence intervals or prediction intervals.

VIII. Discussion and Questions

* Open the floor for questions and discussion.
* Be prepared to answer questions about the process, results, and findings, including any limitations or potential applications of the research.
* Exploratory data analysis: This section involves plotting, analyzing and summarizing the data for time series forecasting. Some of the methods that can be used in this section are:
  + Plotting the data and observing the patterns and trends (e.g. seasonality, cyclicity, trend).
  + Calculating descriptive statistics and summary measures (e.g. mean, median, standard deviation, skewness, kurtosis).
  + Performing correlation analysis and testing for autocorrelation and partial autocorrelation (e.g. ACF, PACF plots).
* Model testing: This section involves choosing, fitting and evaluating different models for time series forecasting. Some of the models that can be used in this section are:
  + [Seasonal Autoregressive Integrated Moving Average (SARIMA): A model that captures both linear and seasonal dependencies in a time series**1**](https://www.advancinganalytics.co.uk/blog/2021/06/22/10-incredibly-useful-time-series-forecasting-algorithms).
  + [Exponential Smoothing: A model that assigns weights to past observations that decay exponentially over time**2**](https://medium.com/analytics-vidhya/7-methods-to-perform-time-series-forecasting-with-python-codes-cc72e72e4e0c).
  + [Long Short-Term Memory (LSTM): A type of recurrent neural network that can learn long-term dependencies in a time series**2**](https://medium.com/analytics-vidhya/7-methods-to-perform-time-series-forecasting-with-python-codes-cc72e72e4e0c).
* Model selection: This section involves comparing, validating and selecting the best model for time series forecasting. Some of the methods that can be used in this section are:
  + Comparing different models or variations of models based on their performance metrics (e.g. MAE, MSE, RMSE) and choosing the best one.
  + Checking for model assumptions and diagnostics (e.g. residuals analysis) and validating the model.
* For exploratory data analysis, some of the methods, tests or plots that can be used are:
  + Plotting the time series and observing its shape, trend, seasonality and outliers (e.g. line plot).
  + Testing for stationarity using statistical tests (e.g. Augmented Dickey-Fuller test) or visual methods (e.g. rolling mean and standard deviation).
  + Decomposing the time series into trend, seasonal and residual components using additive or multiplicative models (e.g. seasonal\_decompose function in Python).
  + Calculating and plotting autocorrelation and partial autocorrelation functions to measure the linear dependence of a time series with its own lagged values (e.g. acf and pacf functions in Python).
  + Detecting change points or structural breaks in a time series using algorithms (e.g. PELT) or visual methods (e.g. CUSUM).
* For model testing, some of the methods, tests or plots that can be used are:
  + Fitting different models or variations of models to the train set using appropriate parameters (e.g. order for SARIMA) and hyperparameters (e.g. learning rate for LSTM).
  + Comparing different models or variations of models based on their performance metrics such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), etc.
  + Checking for model assumptions and diagnostics such as normality, homoscedasticity and independence of residuals using statistical tests (e.g. Jarque-Bera test) or visual methods (e.g. QQ plot).
  + Validating the model using cross-validation techniques such as walk-forward validation or rolling window validation.

Exploratory data analysis:

* Plotting: In addition to plotting the raw stock price data, it may also be useful to plot the returns (percentage change) of the stock price over time, as this can help to identify patterns in the data that may not be immediately apparent when looking at the raw prices. Other useful plots might include histograms of the returns, or boxplots to compare the distribution of returns across different time periods.
* Descriptive statistics: In addition to the basic summary statistics mentioned (mean, median, standard deviation, skewness, kurtosis), other useful measures might include the maximum and minimum values, the range of the data, and the coefficient of variation (CV), which is the ratio of the standard deviation to the mean.
* Correlation analysis: In addition to calculating the Pearson correlation coefficient between the stock prices at different time points, it may also be useful to examine the cross-correlation function (CCF) between the stock prices and other variables that may be related, such as the prices of other stocks in the same industry, or macroeconomic indicators like interest rates or GDP growth.
* Autocorrelation and partial autocorrelation: In addition to examining the ACF and PACF plots to identify potential AR and MA terms for a SARIMA model, it may also be useful to look at higher-order autocorrelations (e.g. ACF and PACF plots for lags > 12 for monthly data) to identify potential seasonality or longer-term dependencies in the data.

Model testing:

* SARIMA: In addition to specifying the appropriate order of the AR, MA, and seasonal components, it may also be necessary to include exogenous variables in the model, such as macroeconomic indicators or news events that could affect the stock price. To select the optimal model, one approach might be to use a grid search over a range of possible parameter values, and choose the model with the lowest AIC or BIC value.
* Exponential smoothing: In addition to the basic exponential smoothing model, other variants that could be tested include Holt's linear exponential smoothing (which includes a trend component) and Holt-Winters' exponential smoothing (which includes both a trend and seasonal component).
* LSTM: In addition to specifying the appropriate architecture and hyperparameters for the LSTM model, it may also be necessary to preprocess the data (e.g. scaling or normalization) and use techniques such as dropout or early stopping to prevent overfitting. To evaluate the model performance, metrics such as mean absolute error (MAE), mean squared error (MSE), or root mean squared error (RMSE) could be used.

For exploratory data analysis:

* Seasonal subseries plots: This involves dividing the time series into seasonal periods and creating a subseries plot for each period to examine the patterns and trends within each season.
* Box plots: This can be used to visualize the distribution of the time series and detect any outliers or extreme values.
* Spectral analysis: This involves decomposing the time series into its frequency components using techniques such as Fourier analysis or wavelet analysis to identify any periodicity or cycles.
* Time series clustering: This involves grouping similar time series together based on their patterns and trends to identify potential similarities or differences between different stocks.

For model testing:

* Bayesian model selection: This involves using Bayesian methods to compare different models or variations of models based on their posterior probabilities, which can provide a more informative measure of model fit compared to traditional model selection techniques.
* Ensemble methods: This involves combining multiple models or variations of models to improve prediction accuracy and reduce overfitting, such as by using bagging, boosting, or stacking techniques.
* Granger causality tests: This involves testing for causal relationships between different stocks or macroeconomic variables that may influence the stock prices, such as by examining the lagged effects of one variable on another.
* Forecast evaluation: This involves comparing the actual forecasted values to the observed values and examining the accuracy and reliability of the forecasts, such as by using metrics such as mean absolute percentage error (MAPE) or symmetric mean absolute percentage error (SMAPE).