APPLICATION SURVEY OF TIME SERIES TREND PREDICTION MODELS ON HOTEL BOOKING DATASET

PRESENTED BY:
SAMBIT MUKHERJEE
LINGJIA SHI

PROBLEM STATEMENT

The tourism industry is one of the most important sectors in the global market

Hotel Booking is a major part of tourism industry.

The accuracy in predicting demands is valued as it is crucial in conducting revenue and resource management for hotels

Choose the best Model to forecast Hotel Booking Trend

Input data samples of hotels booked during every month of a 12-year period

Output the RMSE and MAPE scores in a dataframe and compare them

TECHNICAL CHALLENGES

- To predict a growing trend on a small sample dataset and get the best trend prediction
- To incorporate the changes in seasonality of the trends, e.g. there are sharp ascent in gradient of number of hotels booked during winter or spring months
- Since this dataset looks to have been artificially designed for a trend prediction, how do we see a trend on an actual survey. Because the values might be random increase or decrease in certain months of certain years

RELATED WORKS

- 1990: Box-Jenkins and exponential smoothing, for forecasting hotel occupancy rates
- 2010: Autoregressive (AR) model used to predict intervals for tourist arrivals
- 2011: A combination of linear and nonlinear statistical models used to examine a performance of the forecasting accuracy of demand in Taiwanese outbound tourism
- 2016: A combination of backpropagation and genetic algorithms to predict the foreign tourist arrivals in three cities of Indonesia
- 2017: Combination of biquadratic polynomial function and autoregressive model used for prediction of monthly tourist arrivals in Nepal

SELECTED MODELS

- **Simple Forecasting Methods:**
  - Simple Moving Average Method (SMA**):**

    Forecasts the future value of a time series data using average of the past N observations

$$SMA = \frac{A_1 + A_2 + ... + A_n}{n}$$

SELECTED MODELS

- **Auto Regressive (AR)**
  - Seasonal Auto Regressive Integrated Moving Average **(SARIMA):**

SARIMA method can effectively capture complex relationships in data series as it takes both seasonal and non-seasonal error terms and observations of lagged variables into account when training the model.

Therefore, the model can produce reasonable forecasts which are in line with recent changes in the data series.

$$\varphi_p(B)\phi_p(B^S)\nabla^d \nabla_S^D y_t = \theta_p(B)\Theta_Q(B^S)\tau_t$$

SELECTED MODELS

- **Exponential Smoothing Methods**
  - Holts Method With Trend (**Double Exponential Smoothing):**

    Addresses both the level(l) and trend (b) component of the time series.

$$l(t) = \alpha * y(t) + (1-\alpha)*(l(t-1)+b(t-1)) \text{ — — — — — } \textbf{Level } l$$

$$b(t) = \beta * (l(t) - l(t-1)) + (1-\beta)* b(t-1) \text{ — — — — } \textbf{Trend } b$$

$$y(t+1) = l(t) + b(t) \text{ — — — — } \textbf{Forecast}$$

  - Holts Winter's Multiplicative Method With Trend and Seasonality (**Triple Exponential Smoothing):**

    Apply smoothing to seasonal component in addition to level and trend components.

$$\text{(Level) } L_t = \alpha * (Y_t - S_{t-s}) + (1 - \alpha) * (L_{t-1} + b_{t-1})$$

$$\text{(Trend) } b_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * b_{t-1}$$

$$\text{(Seasonal) } S_t = \gamma * (Y_t - L_t) + (1 - \gamma) * S_{t-s}$$

$$\text{(Forecast for period } m\text{) } F_{t+m} = L_t + m*b_t + S_{t+m-s}$$

TWO METRICS

- **RMSE: The square root of the average of the squared errors.**

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

- **MAPE: Mean Absolute Percentage Error. It is the average of absolute percentage error.**

$$MAPE = \frac{\sum \frac{|A-F|}{A} \times 100}{N}$$

APPROACHES AND RESULTS

OUR DATASET

- Time series dataset consisting of hotel bookings from the year 2010-2022

Month	Bookings
2010-01	120
2010-02	125
2010-03	129
2010-04	126
2010-05	121
2010-06	132
2010-07	150
2010-08	149
2010-09	137

Table 1. Part of our dataset

INTRODUCTION TO EXPERIMENTATION

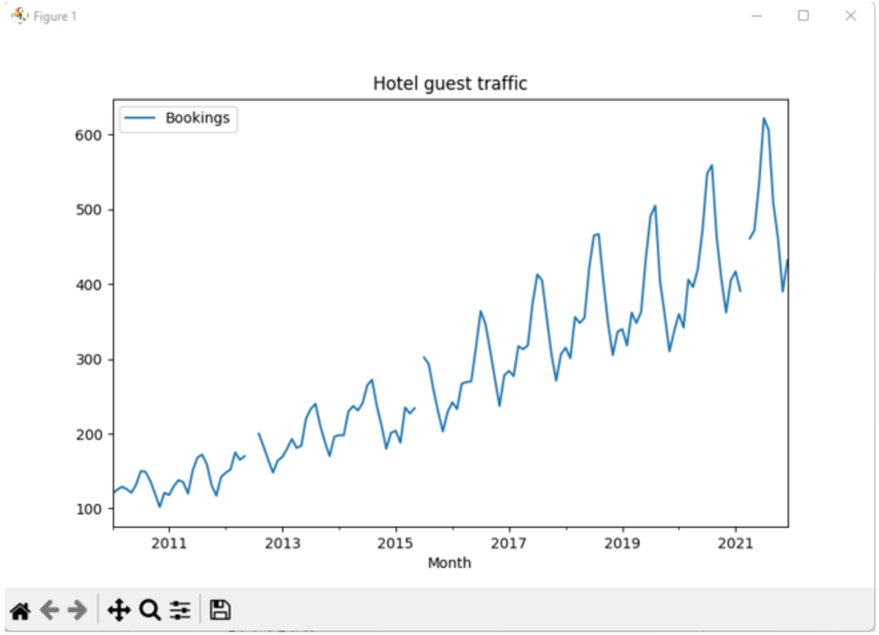
- Our Experimentation includes:
  - Preprocessing the data
  - Building and Fitting the Models to our dataset
  - Plotting the graph for the models
  - Calculating the accuracy of the models through RMSE and MAPE Metrics

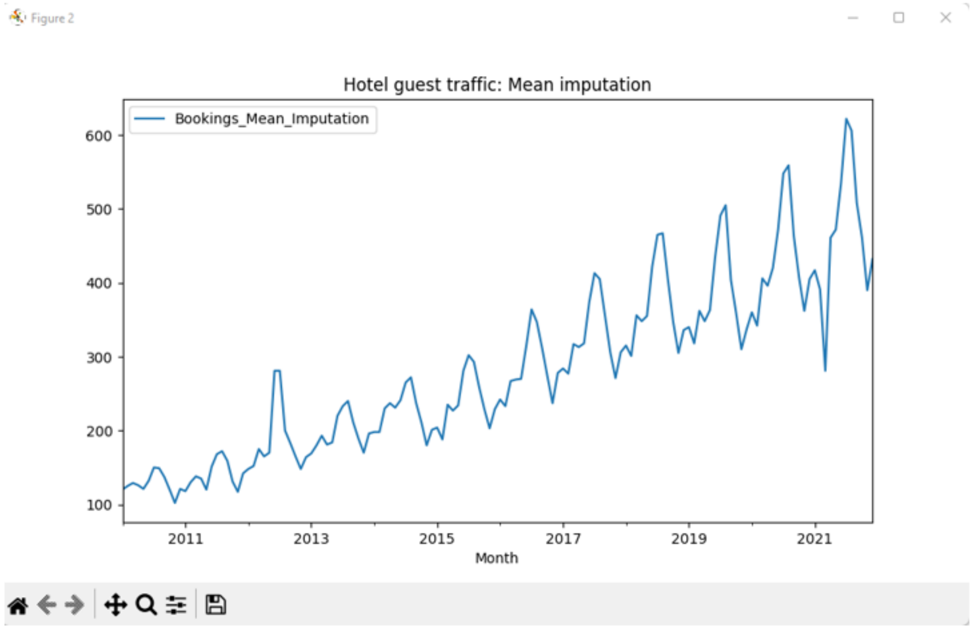
PREPROCESSING THE DATA

- Preprocessing is composed of three steps for our model:
  - Data Imputation
  - Seasonal Decomposition
  - Outlier Detection

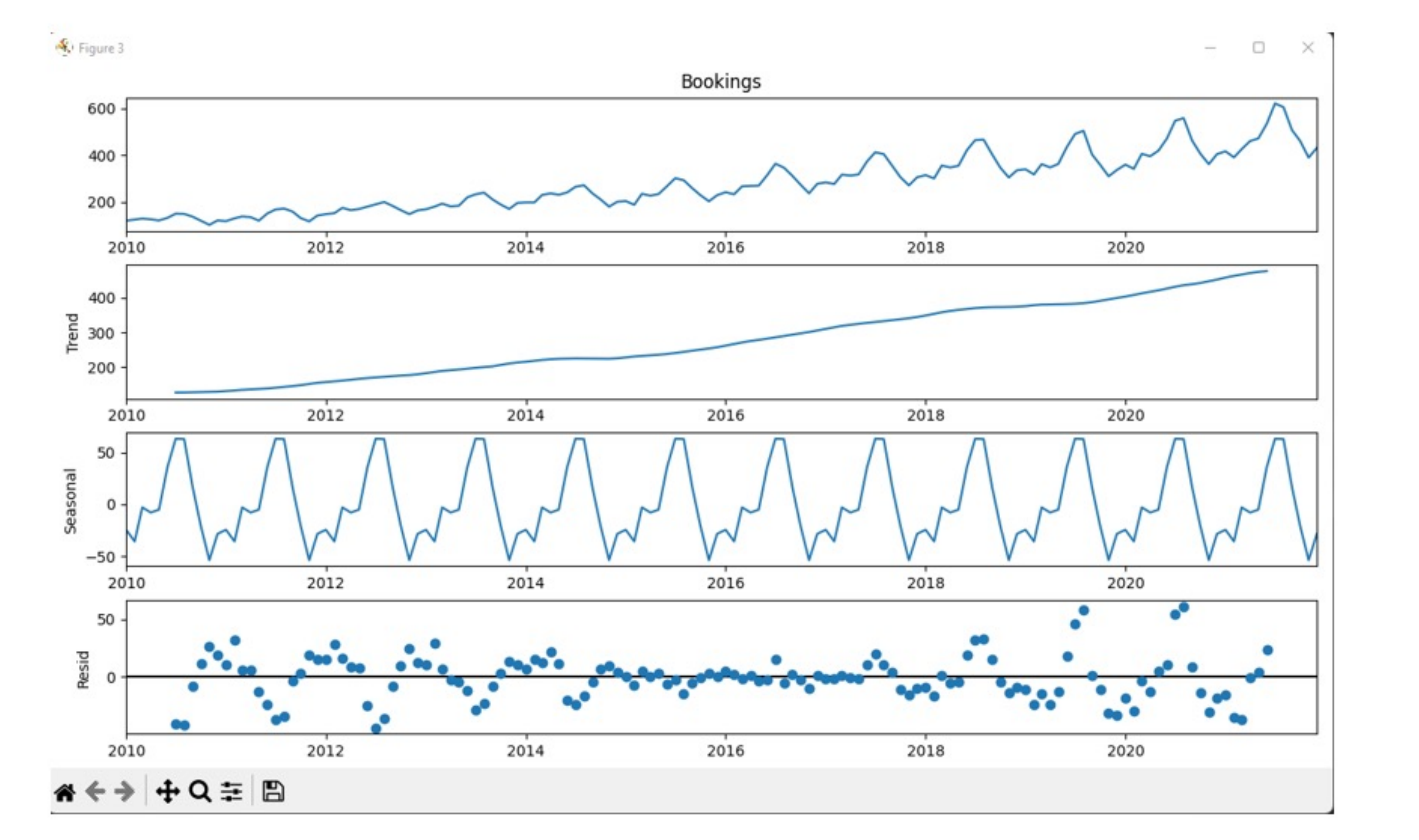
DATA IMPUTATION

- Used to fill the null values in the dataset
- Mean Imputation
- Linear Interpolation

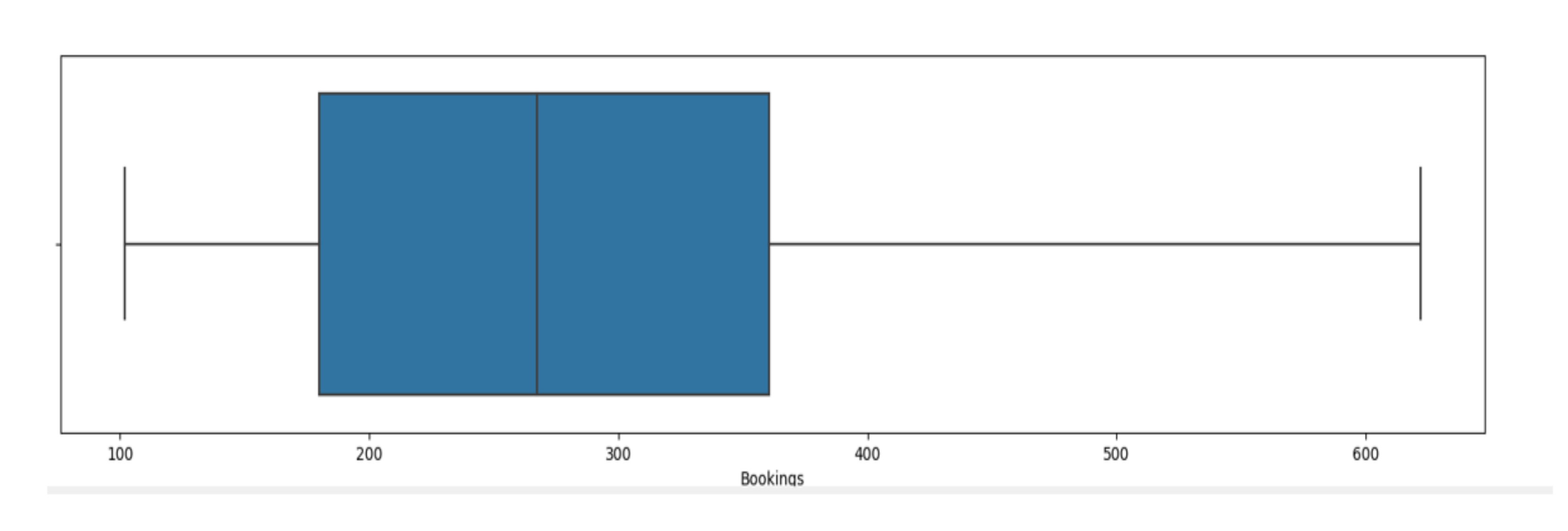




SEASONAL DECOMPOSITION



OUTLIER DETECTION



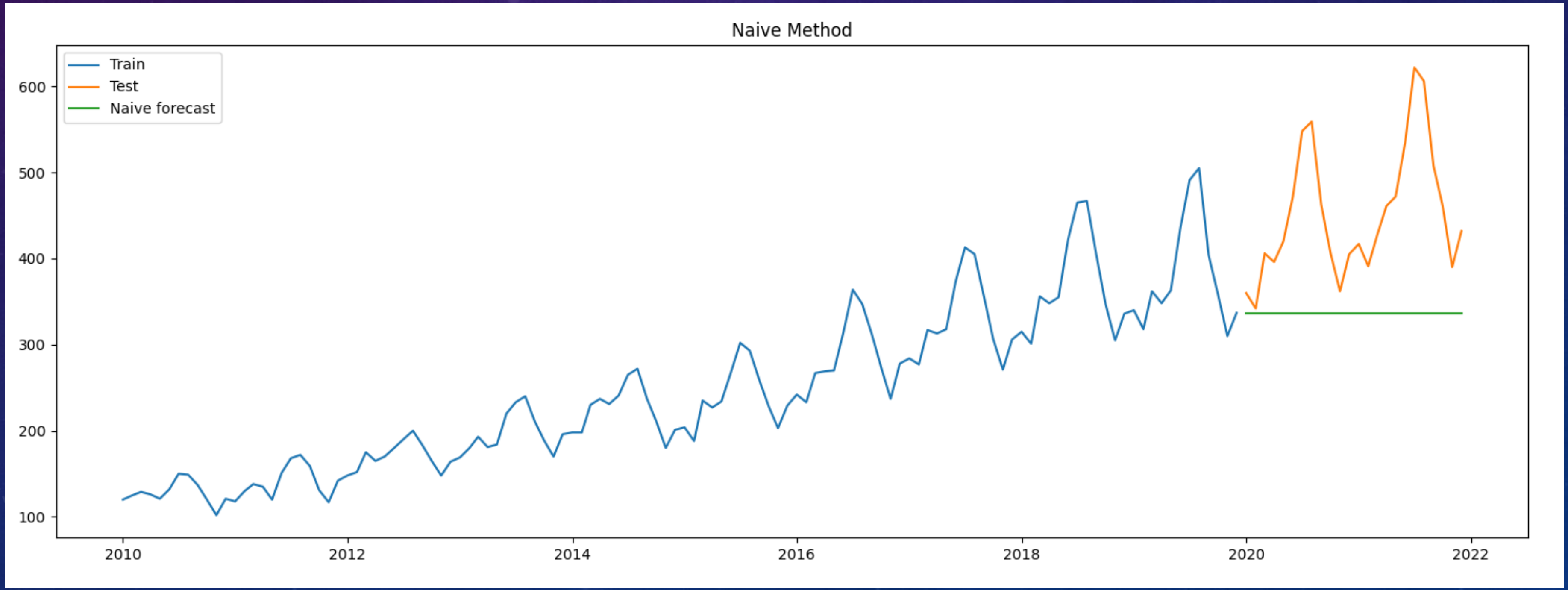
BUILDING THE MODELS

- Used three types of Time series prediction Methods
    - **Simple Forecasting Methods**
    - **ARIMA Methods**
    - **Exponential Smoothing Methods**

SIMPLE FORECASTING METHODS

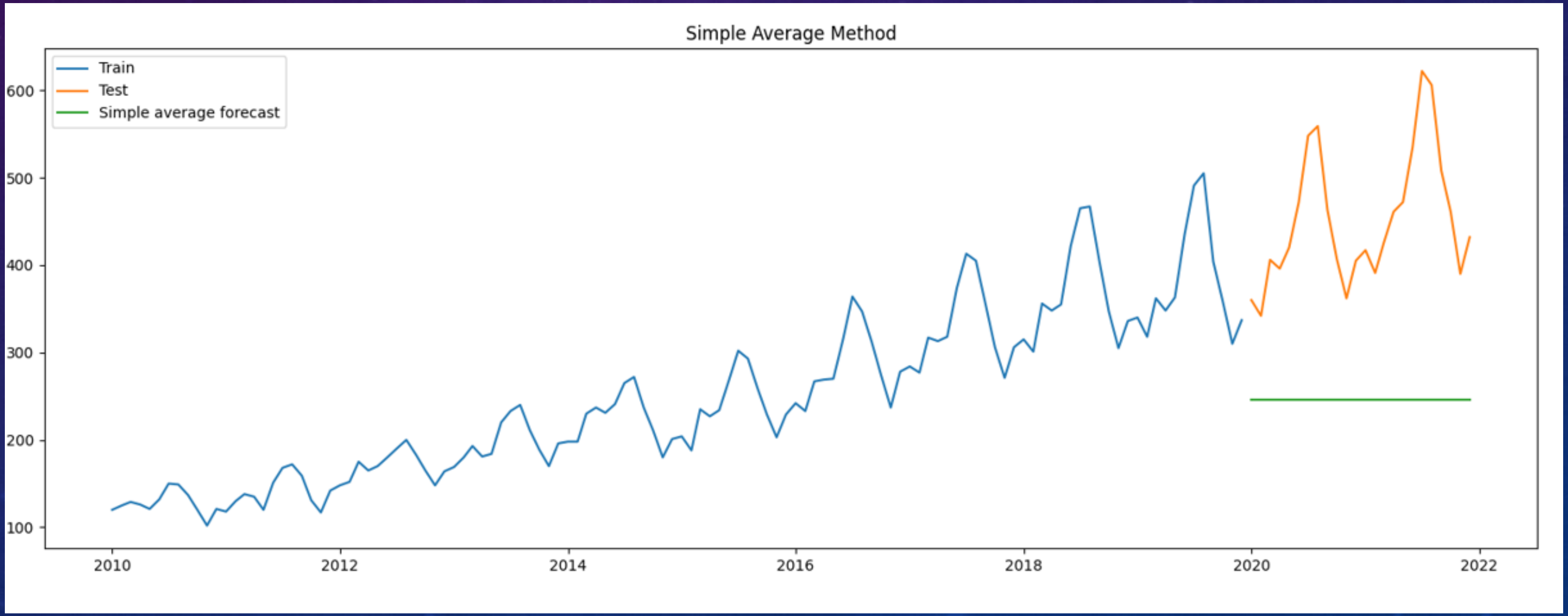
NAÏVE MODEL (NM)

- No handling of causal relationships
- Indicated by the forecast line(green line)
- Forecast does not pas through any data points



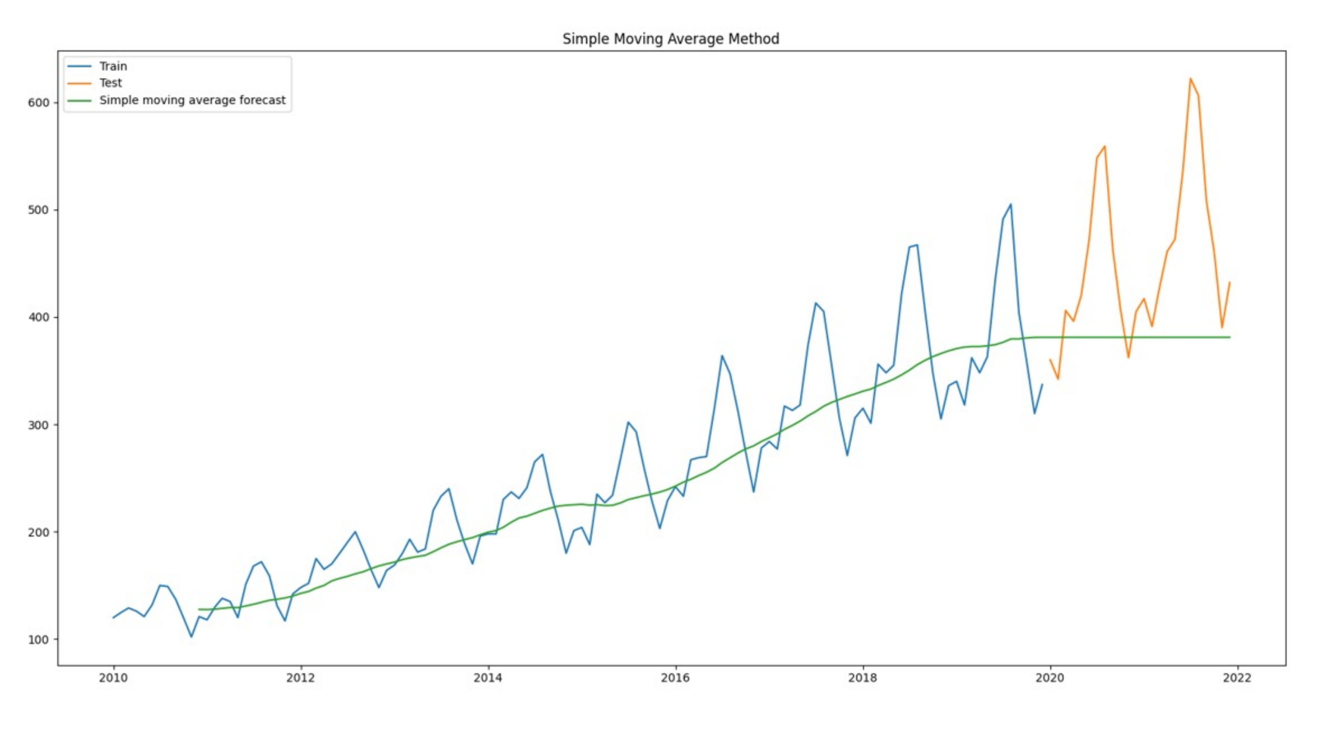
SIMPLE AVERAGE METHOD (SAM)

- Does not consider outliers
- And thus heavily affected by outliers



SIMPLE MOVING AVERAGE METHOD (SMAM)

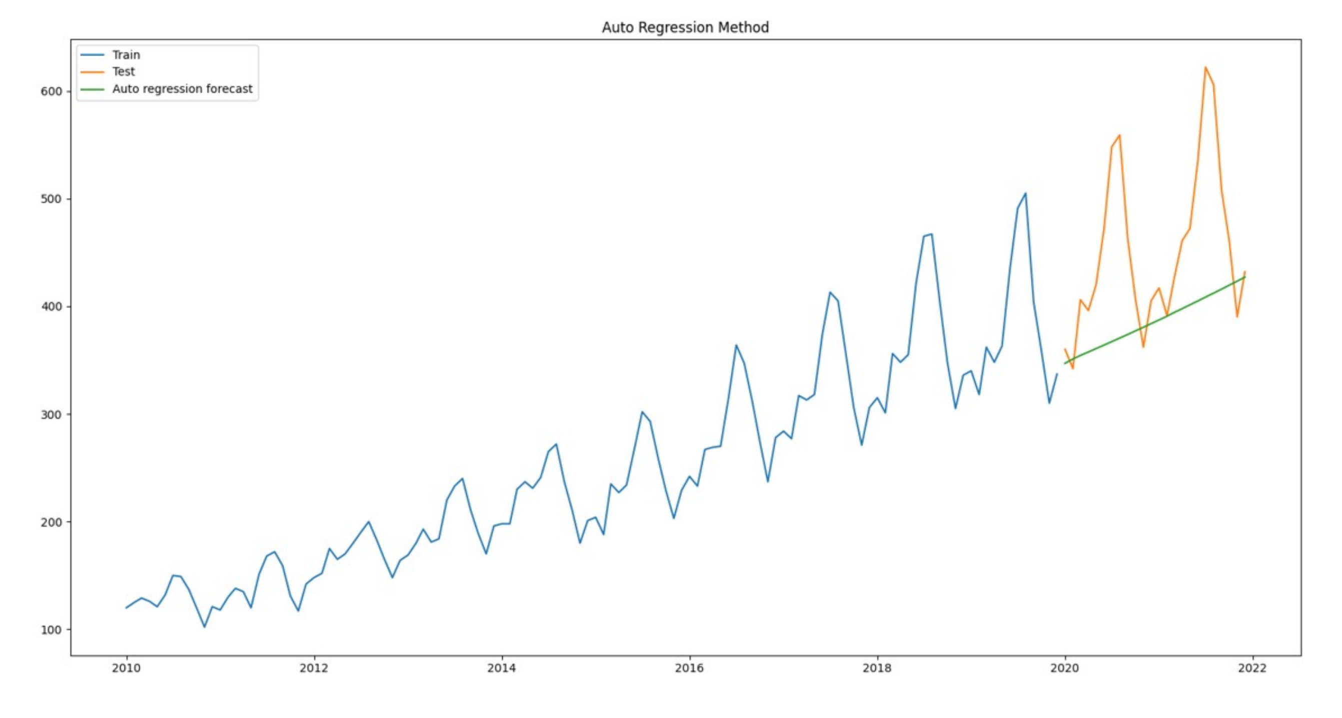
- Window period of 12 years
- Forecasts the results more accurately than both our previous models
- The accuracy of forecast on training set is much higher than the test set



ARIMA MODELS

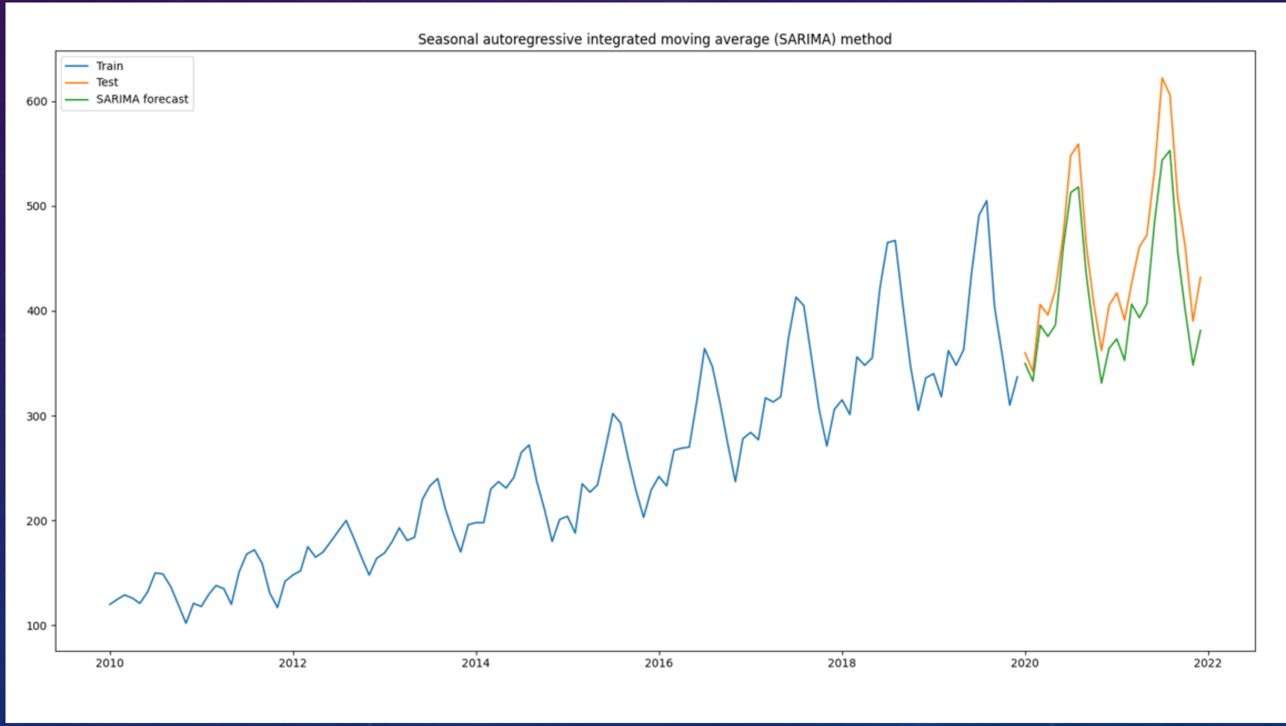
AUTOREGRESSIVE METHODS (AR)

- Good trend predicted
- Does not consider shifting trend
- Predicts a linearly increasing trend



SEASONAL AUTO REGRESSIVE INTEGRATED MOVING AVERAGE (SARIMA)

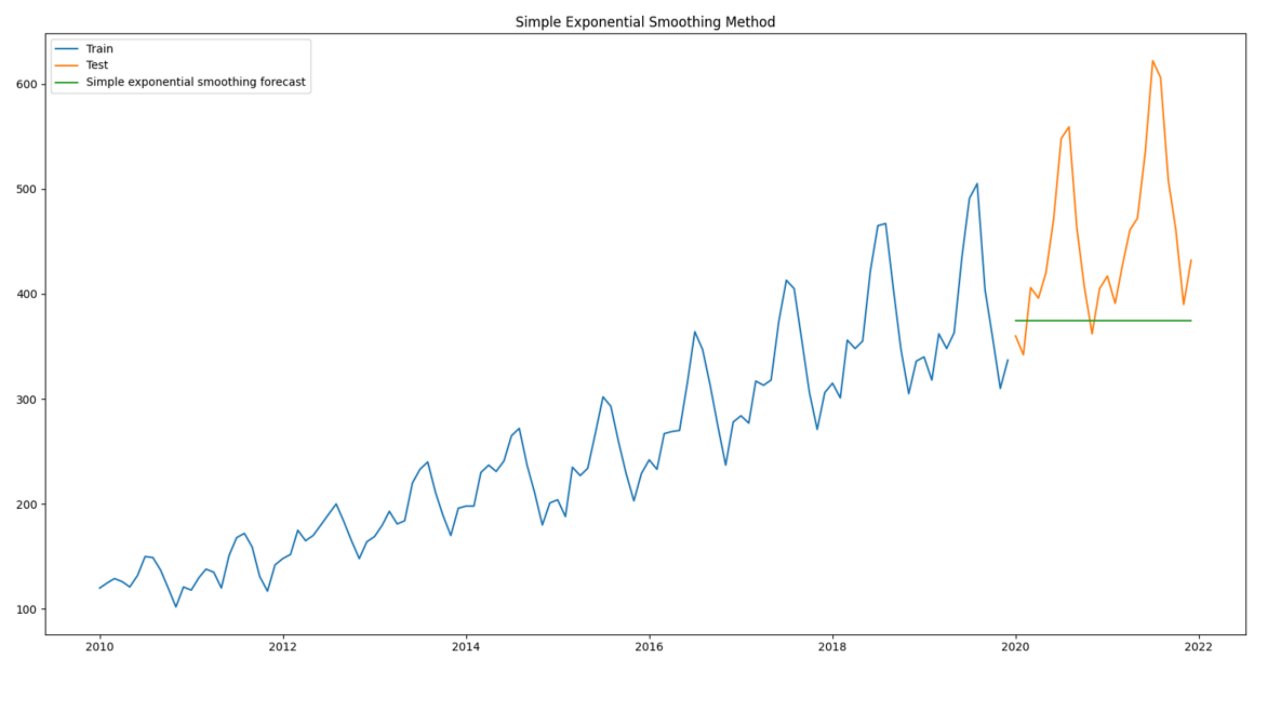
- Considers positive and negative shift in trend and seasonality over time
- Second Best Model



EXPONENTIAL SMOOTHING METHODS

SIMPLE EXPONENTIAL SMOOTHING METHOD (SESM)

- Actually a good method of prediction
- Dataset is relatively small, therefore full results not showing
- Larger dataset is likely to have a better result.



HOLTS METHOD WITH TREND (HMWT)

- A trend begins to emerge immediately even on a smaller set
- Seasonal period of 12 months
- Additive Trend Used

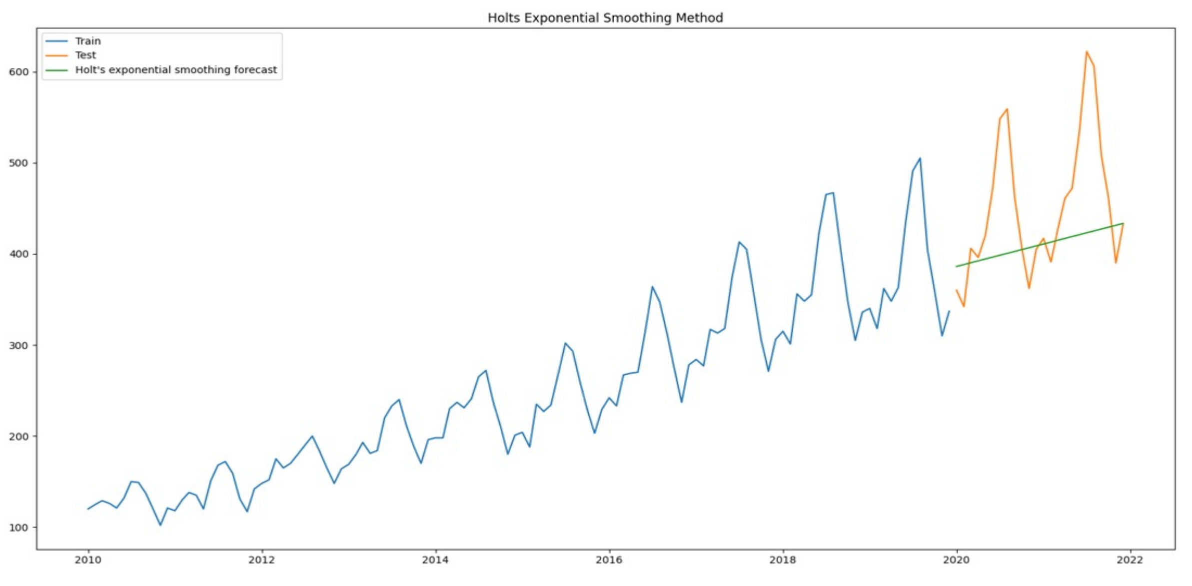
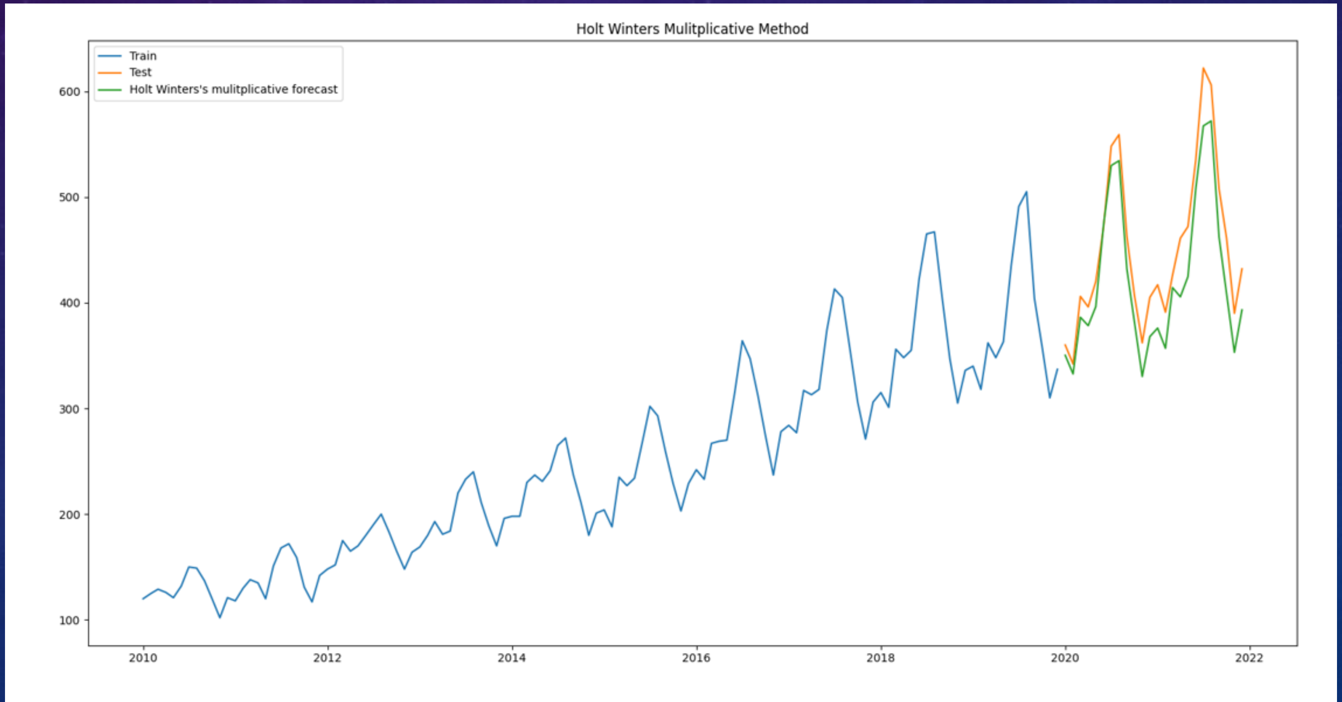


Figure 8: Holts linear trend forecast

HOLTS WINTER'S MULTIPLICATIVE METHOD WITH TREND AND SEASONALITY (HWMM)

- Best result
- Very close approximation of the shifting seasonality
- Best for the datasets, where the trend and seasonality are increasing over time.



RESULTS

Table 2. RMSE and MAPE scores of different methods

Method	RMSE	MAPE
NM	137.51	23.63
SAM	219.40	44.23
SMAM	103.33	15.54
AR	95.71	14.18
SARIMA	42.86	8.39
SESM	107.65	16.49
HESM	82.31	11.57
HWMM	33.72	6.71

CONCLUSION

- No trend prediction and no seasonality variance accounting gives bad results
- Trend prediction, no seasonality change, gives average results
- Trend prediction with Seasonality gives best results

BROADER IMPACT OF TIME SERIES TREND PREDICTION

- Demand forecasting for retail, procurement, and dynamic pricing
- Price prediction for customer-facing apps and better user experience
- Forecasting pandemic spread, diagnosis, and medication planning in healthcare
- Anomaly detection for fraud detection

THANK YOU