**APPLIED TIME SERIES ANALYSIS**

**FINAL PROJECT: RAINFALL FORECASTING**

S042-Vikramaditya Singhai

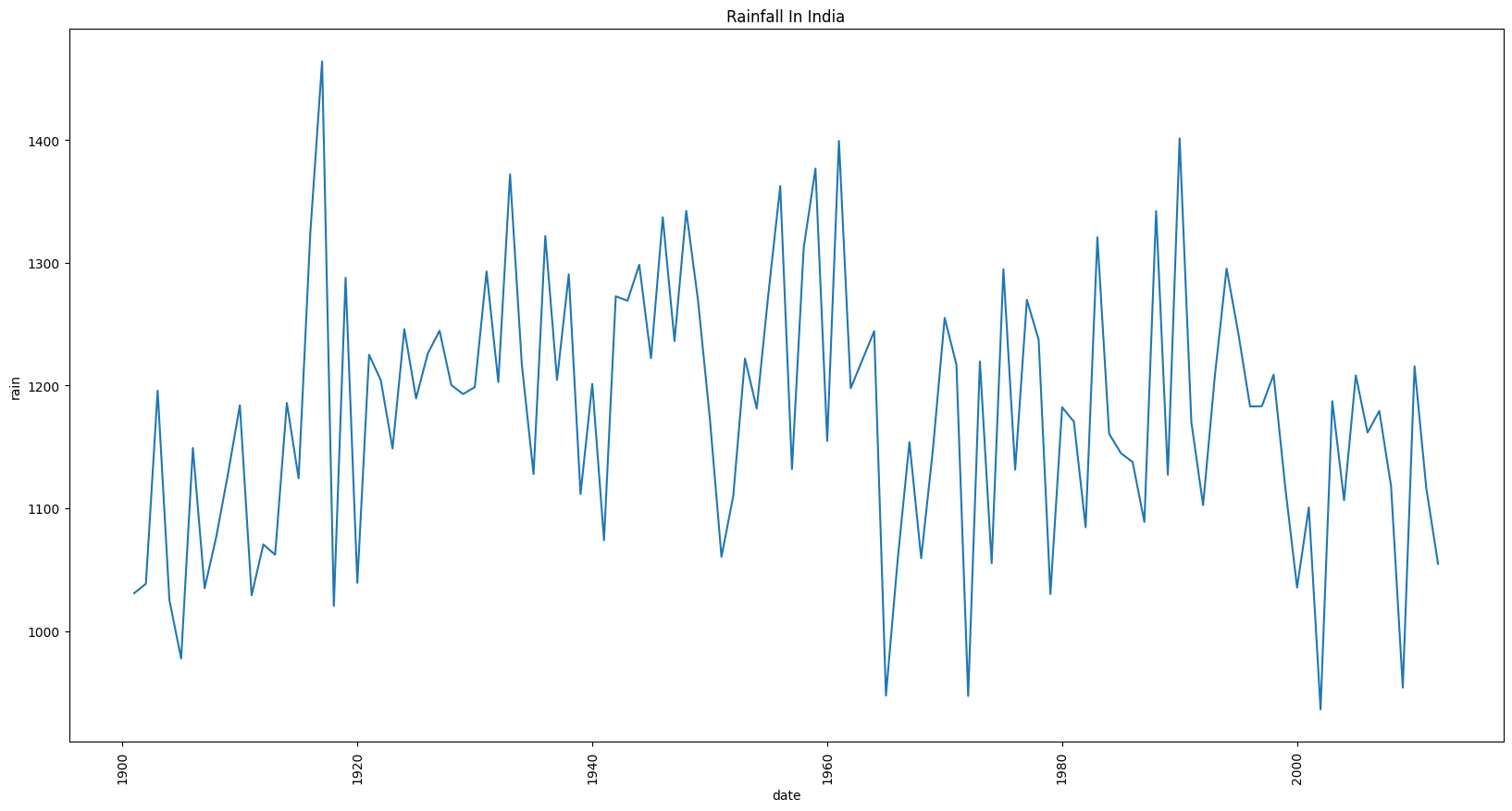
S043- Arina Singhai

**SUMMARY**

This time series modeling project involved analyzing historical rain data (1901-2012) to identify patterns like seasonality and autocorrelation to fit a statistical forecasting model. After exploratory data analysis and preprocessing, model selection and hyperparameter tuning were used to optimize the parameters. The fitted model was trained on the available data, incorporating key data characteristics into the forecasts. Model performance was evaluated by generating predictions on withheld validation data and comparing errors against benchmarks. This project demonstrates a generalizable workflow for time series forecasting, from exploratory analysis to model validation.

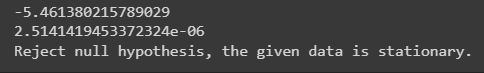
**MODELLING PROCEDURE**

1. **Visualising the data**

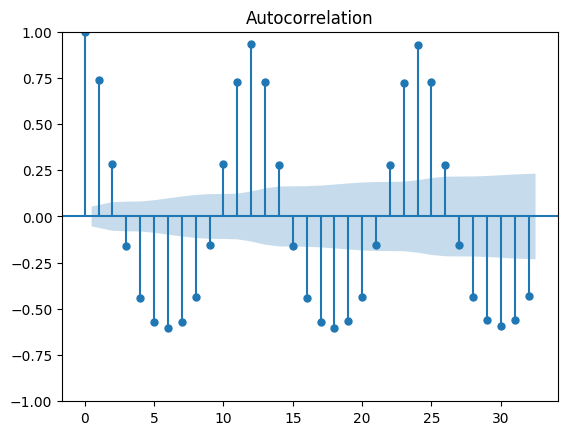
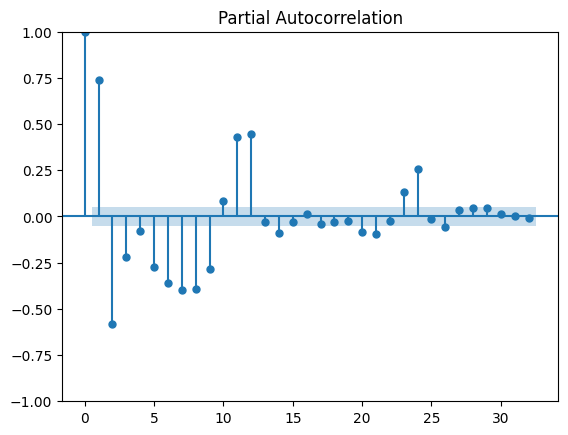


1. On checking the data, we found no unusual observations, and no transformational techniques were needed to be applied.
2. **Checking Stationarity of the data**

* The Augmented Dickey-Fuller test is used to check if a time series is stationary or not.
* It tests the null hypothesis that the time series has a unit root, meaning it is non-stationary. The alternative hypothesis is that the series is stationary.
* A small p-value (< 0.05 for example) means the t-statistic is far from zero in the negative direction. This implies strong evidence against the null, meaning the series appears stationary.
* A large p-value (> 0.05) implies we cannot reject the null hypothesis that there is a unit root at the chosen significance level. The t-statistic is not sufficiently negative to suggest stationarity.
* For the given Annual Rainfall data:



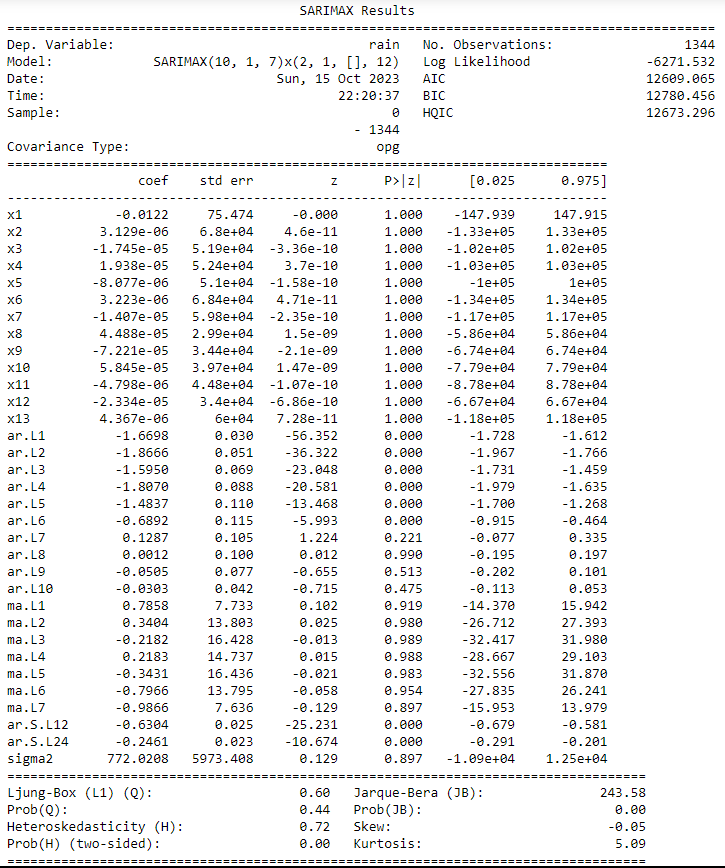
1. **ACF AND PACF GRAPH**



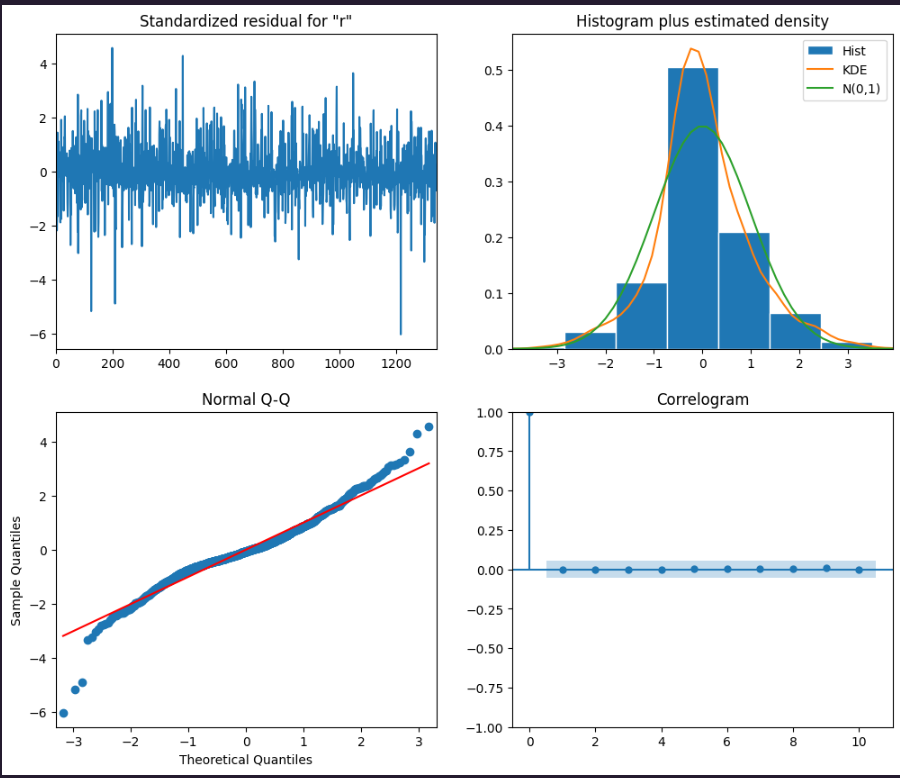
ACF is used to compute the correlation of the time series with its lag. Here The ACF shows a seasonal pattern that keeps repeating after certain lag values ( here 12 ) showing a seasonal pattern followed in the data.

PACF measures the **unique** correlation between a time series and its lagged values, after accounting for the correlation with other lagged values.

1. **BEST MODEL**

****

This Is The Current Best Model With the order of (10, 1, 7) – 10 components for the AR Model, 1 differencing done and 7 For the MA Model, Along with a seasonal P of 2 and a seasonal repetition of every 12 months along with the exogenous data while giving the graph the lowest AIC and BIC possible.

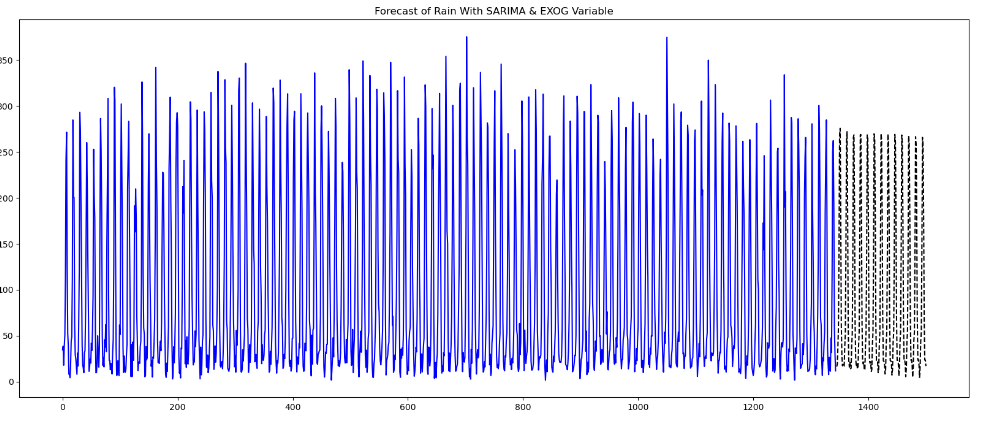
****

Here, the Standardized residuals graph shows that the residuals have a mean of 0 and a variance of 1. Residuals here refer to the differences between the observed values of the time series and the values predicted by the fitted model.

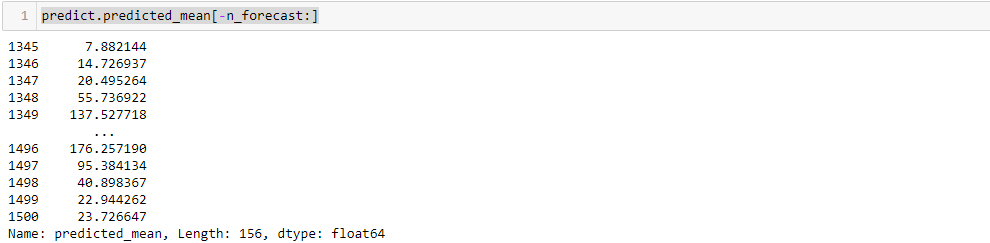
A Hist Plot With Estimated Density is a plot that shows a histogram of the standardized residuals, along with a density estimate. It is a smooth curve that represents the probability distribution of the standardized residuals and if it’s a bell-shaped curve, it suggests that the standardized residuals are normally distributed.

Q-Q Plot shows the plot of the quantiles of the standardized residuals to the quantiles of a normal distribution. If the standardized residuals are normally distributed, then the points on the Q-Q plot should fall close to the diagonal line.

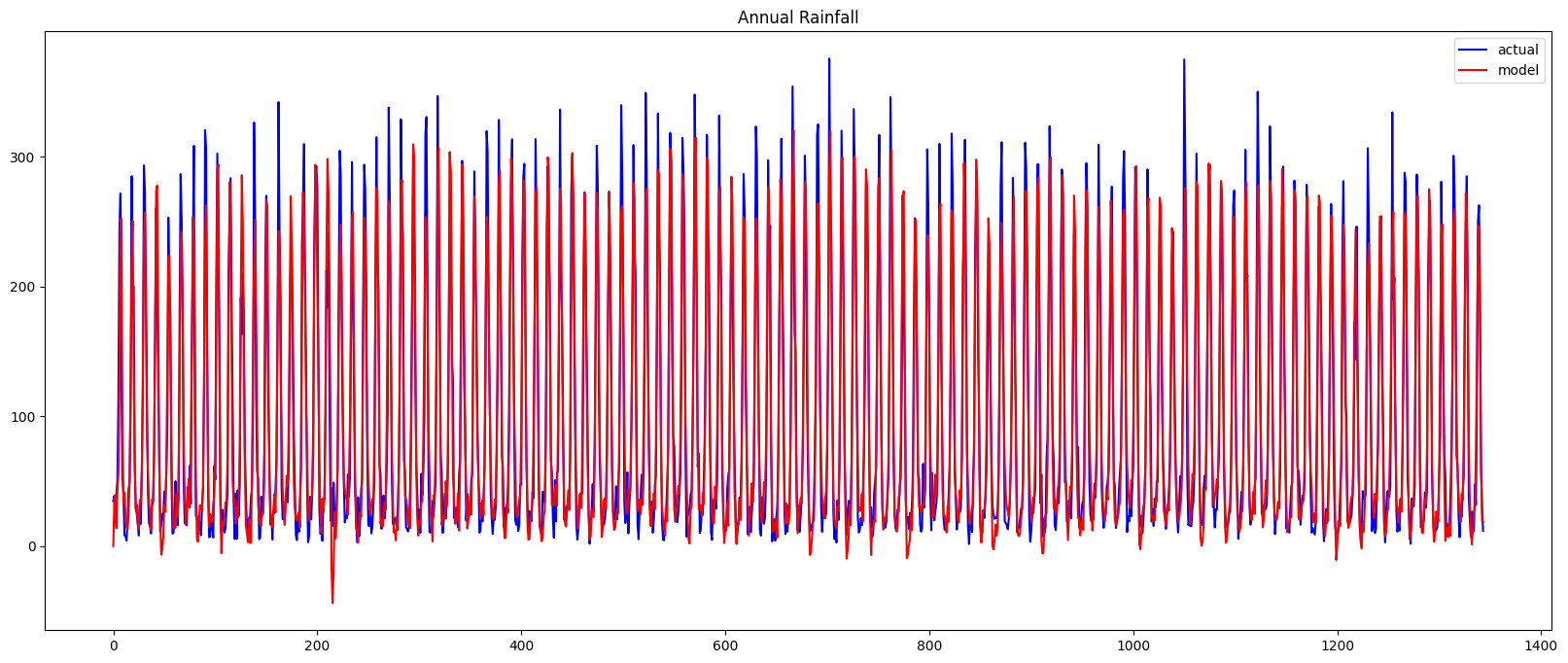
The correlogram is just the ACF and PACF of the standardized residuals where the ACF and PACF should just be as close to 0 as possible except Lag 1.



This Graph shows the best predictions possible with the SARIMAX Model + The Exogenous Variables. As seen, there is a slight downward trend to the predicted values.



Predicted Values

  
This graph shows the model predictions of our model as compared to the actual data and we see a good model prediction as compared to the data.

**CONCLUSION:**

We used a SARIMAX model to forecast annual rainfall in India. The model captures the complex seasonal and autocorrelation patterns in the data. We found the best model parameters by exploring the data and tuning the hyperparameters. The model accuracy and diagnostics show that our methodology is sound. Overall, this project demonstrates a robust time series analysis workflow that could be used to forecast other seasonal climate variables. The techniques we used are a data-driven way to generate reliable forecasts from historical datasets.