|  |  |
| --- | --- |
| Name of the Student | Priyank D Gala |
| Internship Project Title | Air Passegers Time Series Model |
| Name of the Company | TCS ion |
| Name of Industry Mentor | Esmita Gupta |
| Name of the Institute | BK Birla College of Arts, Science & Commerce |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort | Project Environment | Tools Used |
| 19th Oct 2022 | 19th Nov 2022 | 78 hrs | Google Colab | Python |

**Table of Contents**

[Acknowledgements 2](#_Toc120032252)

[Objective 3](#_Toc120032253)

[Introduction 4](#_Toc120032254)

[Internship Activities 5](#_Toc120032255)

[Methodology 6](#_Toc120032256)

[Assumptions 6](#_Toc120032257)

[Exclusions 6](#_Toc120032258)

[Charts 7](#_Toc120032259)

[Algorithms 14](#_Toc120032260)

[Challenges & Opportunities 14](#_Toc120032261)

[Risk vs Rewards 14](#_Toc120032262)

[Reflections on the Internship 14](#_Toc120032263)

[Recommendations 14](#_Toc120032264)

[Conclusion 15](#_Toc120032265)

[Enhancement Scope 15](#_Toc120032266)

[Link to code and executable file 15](#_Toc120032267)

## Acknowledgements

I would like to express my gratitude and special thanks to Vice Principal and Head of the Department of IT Mrs. Esmita Gupta ma’am for guidance in support throughout in my project to improve my skills and get better results.

I would also like to thank TCS ion for giving me a platform in opportunity in this internship to learn the deep learning machine learning and new model making this has helped me to improve and embed new skills.

DATE:

19th November 2022

## Objective

The objective of this Internship Project is to find out patterns and predict weather the passengers will by using aeroplane for travelling or not based on the historical data and time series model of ARIMA and SARIMAX. Also, to find out that how much are model is efficient and accurate and forecasting the results.

## Introduction

For our project we have chosen the airplane dataset from the Kaggle in the CSV format. In our data sets we have the dates and passengers on board on the date so we have these two columns which will help us to find that is there any pattern by which people are travelling and predict the future travelling numbers.

## Internship Activities

Act 1:

* Read the dataset and understood the dataset.
* Got info about all columns in Dataset and their types.
* Applying Discrete Statistics.

Act 2:

* Creating a sample dataset to train and use our base model.
* Grouping and sorting by the dates.
* Using Seaborn to forecast the passengers travelling by year.

Act 3:

* Forecasting bar graphs for every month of every year the passengers have travelled.
* Creating a line graph to check the growth of passengers over the years.
* Forecasting stationery statistics for mean, standard deviation, p-Value, critical value, etc.

Act 4:

* Forecasting the model to find seasonal trend and to Verify the P value.
* Creating a model to forecast the moving average of a data.
* Creating a model to find out the lock transformation of her model versus the original.

Act 5:

* Creating an auto ARIMA model for a training dataset and forecasting the prediction results.
* Finding out the autocorrelation and the ARIMA value for our model.
* Forecasting the autocorrelation and partial correlation in our model.

Act 6:

* Building a SARIMAX model based on our ARIMA results.
* Finding out the mean square error, root mean square error, mean absolute error, and mean absolute percentage error.
* Forecasting the final results of our SARIMAX model for prediction.

## Methodology

Our method was to find the pattern in the increment of passengers over the year and, after we did that, we have to apply discrete statistics to get a statistical report on how the passenger trends are going. We found our p-value and other statistical results giving r model the accuracy of 95% for the prediction.

## Assumptions

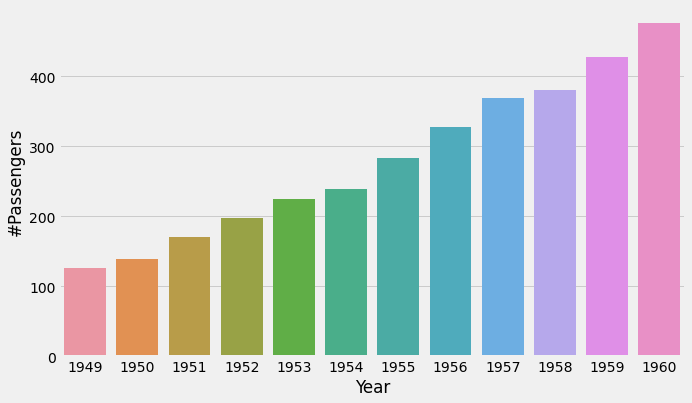
The assumptions were there will be increment in the number of passengers using the airplane. And the passengers over the years will be choosing airplane for travelling over other vehicles.

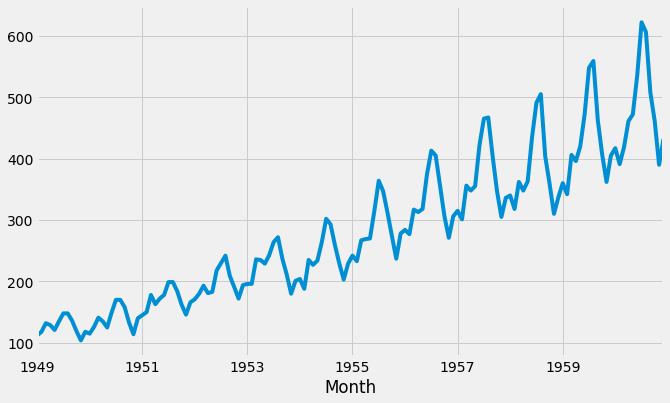
## Exclusions

We excluded every other vehicle and just focused on a timeline as how the passengers have been travelling through the aeroplane and how they will be more travelling with the air services in the upcoming future.

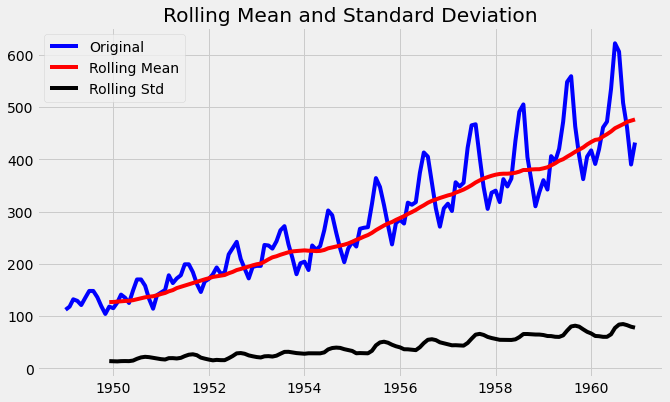
## Charts

Passengers Travelling Over the Year:





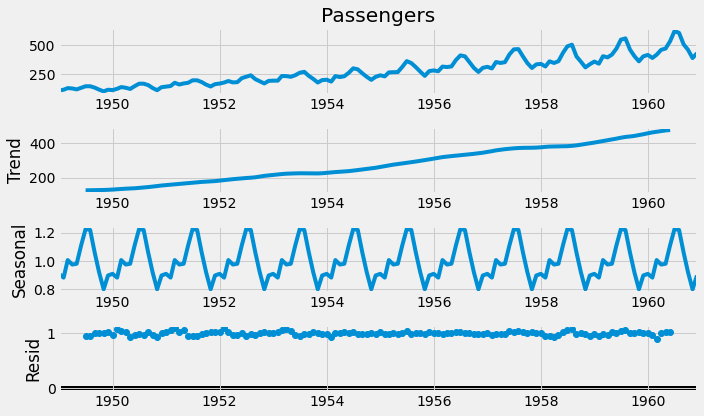
Discrete Stats:



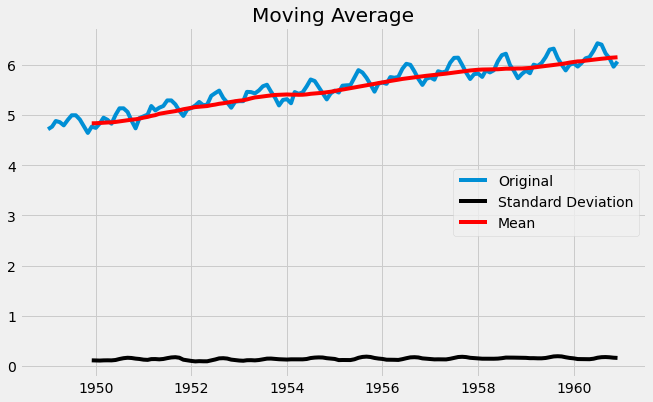
Results of Discrete Stats:

|  |  |
| --- | --- |
| Test Statistics | 0.815369 |
| p-value | 0.99188 |
| No. of lags used | 13.0000 |
| No. of Observations used | 130.000 |
| Critical Value (1%) | -3.48168 |
| Critical Value (5%) | -2.88404 |
| Critical Value (10%) | -2.57877 |

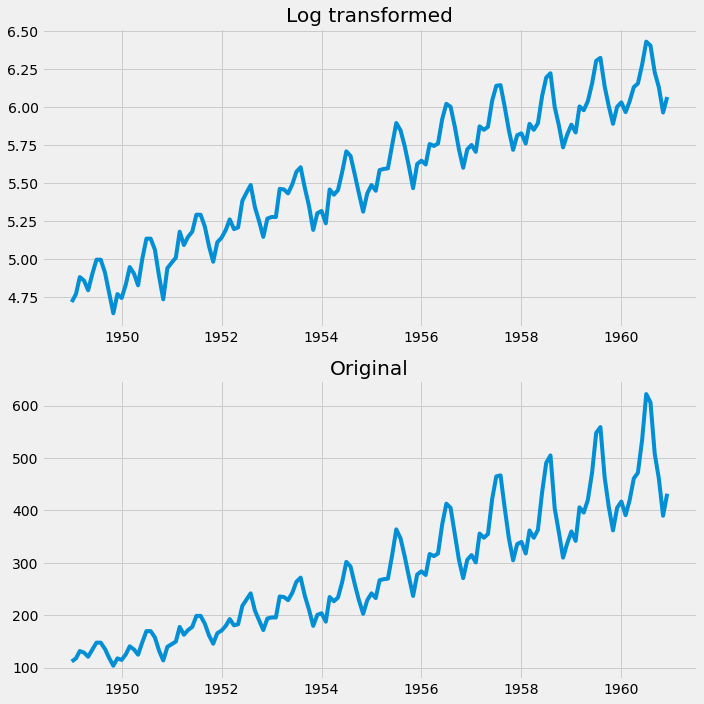
Trend and Seasonal:



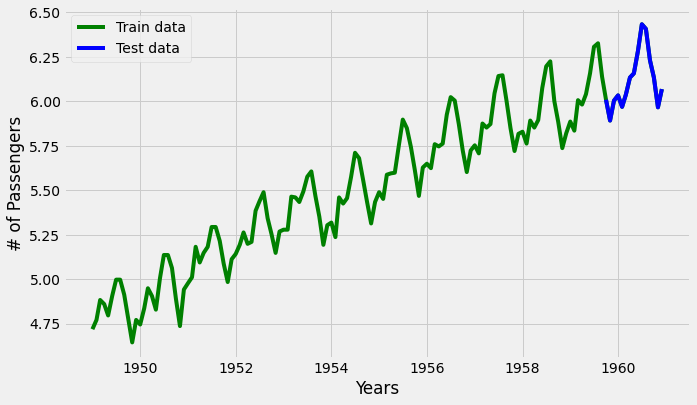
Moving Average of Passengers:



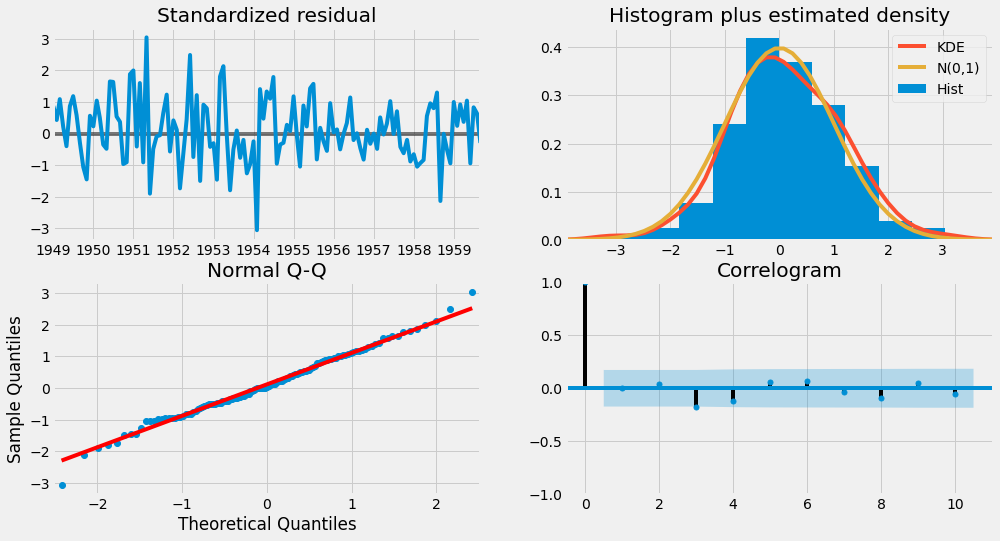
Log Transformed v Original



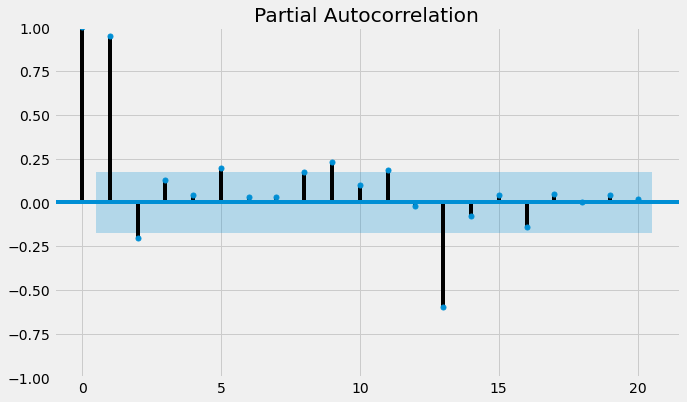
Test v Train Data in ARIMA Model

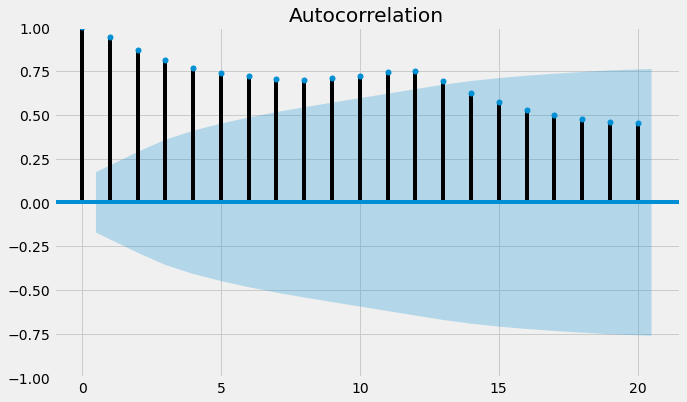


ARIMA Results:

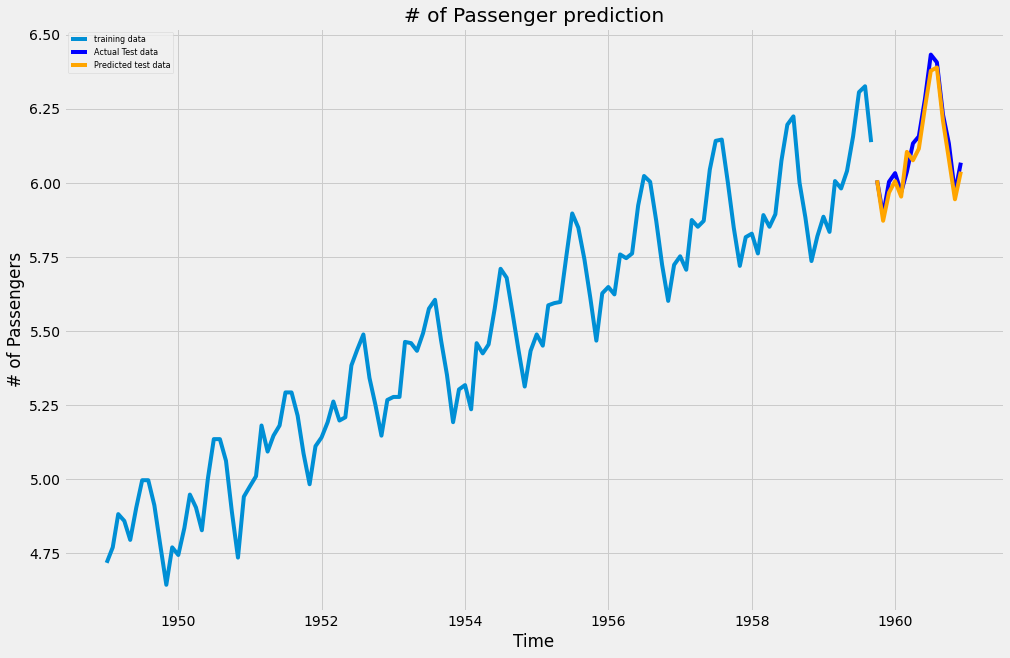


Correlations:





Final Prediction Forecast



## Algorithms

We used our ARIMA algorithm and discrete mathematics to find the perfect accurate number for our SARIMAX model. After that we created a test data to predict the future values and then we forecasted the results.

Then we used are SARIMAX model with P-value and all are discrete statistics, to get a forecast of our predicted values of the future of air passengers travelling.

## Challenges & Opportunities

The first challenge in a dataset was there no other columns were there it was solely dependent on the time series on how passengers are travelling then we need to find out if any pattern is there or not and we didn’t found any patterns we have to only rely on our discrete statistics we used our hypothesis and P-value we got the 95% accurate results so this was very challenging opportunity to find out whether there are any trains or not without other dependencies on data.

## Risk vs Rewards

There is cause there was nothing data dependable column but even with that with the magic of time series we got a predictions with 95% accurate as our reward.

## Reflections on the Internship

It was really tough to build a model for such kind of dataset where we don’t have any other dependencies and we have to predict using the time series but it was actually fun to do so and to get better results to improve on our best skills and to gather more values India machine learning platform.

## Recommendations

The first recommendation will be that if anybody want to learn the time series then this dataset is easy to understand and to learn and predict the results using the discrete statistics and the ARIMA and SARIMAX models to get results and proper understanding of how these things actually work.

## Conclusion

We were successfully able to predict the number of air passengers will there be an increment or decrement using the historical time series data and observation was there should be an increase in the passengers travelling through air and prediction was the same so with a 95% accuracy offer hypothesis we were successfully able to predict then the passengers will be more likely able to travel through air.

## Enhancement Scope

In a dataset we were only able to find out whether the passengers will be travelling through monthly order or not if we got more data on the seasons and festivals than a model would have been even stronger to predict of each and every month through the help of time so that would be a big enhancement on our data but a data lacked that but still without that we got a very good amount of accuracy and prediction of our data.

## Link to code and executable file

[Click Me](https://colab.research.google.com/drive/18IZmc144wSzhWLoO_kMmliuaN3v6QDji?usp=share_link)