Project Report for Problem 1:

Members: Aditi Jaiswal

Amit Yadav

Abhijit Khodade

Problem Statement: To predict the passenger revenue miles for both the domestic and international air carriers for a period of 20 years. (From 2000 Jan- 2020 March).

Data Collection: The data has been downloaded from the given link <https://www.transtats.bts.gov/TRAFFIC/>

Visualization of the Data: We can see little bit trend present in the dataset and Seasonality present.

Checking if Data is stationary or Not: weak evidence against null hypothesis, time series has a unit root, indicating it is non-stationary.

Checking which components are present- using decomposition method

Decomposition Methods

•Decomposition methods are based on an analysis of the individual components of a time series. The strength of each component is estimated separately and then substituted into a model that explains the behaviour of the time series.

We have two types of Decomposition methods-

1. Additive: xt = Trend + Seasonal +Cyclic+ Random, The multiplicative decomposition model is expressed as the product of the four components of a time series: Yt = TRt × St × Ct × It and Is useful when seasonal variation is increasing overtime
2. Multiplicative: xt = Trend \* Seasonal\* Cyclic \* Random Another approach is to treat each observation of a time series as the sum of these four components. Yt = TRt + St + Ct + It and Is useful when season variation is relatively constant over time.

### Analysis of both the above decomposition methods

From plot we can see there is seasonality present, but also seems trend is there

#### We have few options here; we can use Smoothing techniques to forecast or Make it stationary and Build an ARIMA and SARIMA MODEL

#### We will use Smoothing techniques for now- But later Make it Stationary and Train ARIMA AND SARIMA models also.

Smoothing technique -

Smoothing techniques remove fluctuations in a time series. The fluctuations may be due to seasonal and irregular components, so the result of removing them from the time series reflects only the trend and cyclical components

Types of Smoothing techniques -

1.Moving average

1. Centered Moving average
2. Exponential smoothing technique

Types of Exponential smoothing technique

If no trend or Seasonality present, Use Simple Exponential smoothing technique If trend is present, Use Holt's Trend corrected smoothing technique If both trend and seasonal components present, use Holt's Winter method

Here We saw while decomposing the components that there was trend present and Seasonality are present there Hence Will use Holt's winter method but Will use other methods.

# ARIMA Modelling: Making data stationary before modelling it, Will take differencing of 12 because there is seasonality present in the dataset.

We had performed Ad fuller dickey test again.

#### Ad fuller Test result analysis-

Now P Value is less than 0.05 hence Data is stationary now and Will reject null hypothesis.

# ACF and PACF

ARIMA Model has three parameters p, d, q where p represents AR, d represents differencing and q represents Moving average

And Indentification of AR model is done with PACF , Indentification of MA model is done with ACF.

Test Analysis: Q = 5, 4, 3 , P =2, 1, d is usually between zero to 2 hence will experiment with values between zero to 2 –

We compared Arima method with Exponential method and saw that ARIMA is performing

worse than Exponential method. Thus, we Performed Hyper Parameter tuning for ARIMA.

And Fitted the SARIMA model.

Result from every Model:

Holt's trend method, RMSE = 10621491.869534679

Holt's winter Method ,RMSE= 13876638.261904823

ARIMA Model ,(2,1,3) RMSE = 1868585.2575958828

ARIMA Model , (2,1,4) RMSE= 78579925.94675155

SARIMA model, RMSE = 1383431.9123703397

SARIMA model performed better than all others

Predicted values for next 12 month by SARIMA

218 2.438847e+06

219 2.241160e+06

220 3.015368e+06

221 2.433880e+06

222 2.662910e+06

223 2.702916e+06

224 3.498654e+06

225 3.289709e+06

226 2.537229e+06

227 4.034587e+06

228 3.469450e+06

229 3.262626e+06

Answer to the Questions -

1. How the data changes from month to month and year to year?

– Any trend? - Yes There is little trend present

– How fluctuated the curve is?

– Any seasonal effects? - Yes, Seasonality is present

– Any unusual years/months which have significantly small or large number? - No, There is No such significance large or small values

here is the statistics of data

Total

count 2.420000e+02

mean 6.910944e+07

std 1.154640e+07

min 3.860172e+07

25% 6.153300e+07

50% 6.828799e+07

75% 7.675590e+07

max 1.022825e+08