A project report on

Air Traffic Demand Analysis and Forecasting

Submitted in partial fulfillment of the requirements for the Post Graduate Program in Data Science and Engineering

Submitted by

Aswath Balaji	(SSID: WJNGJDUI90)
Aswathi KV	(SSID: INCRBWMGVL)
Juhi Badiyani	(SSID: 4BXKBW4VF0)
Smaran M Sogi	(SSID: CNY7GZTUP6)
Sulaiman Chandali Kazi	(SSID: HWJUSS0BHI)

Batch: DSE_BLR_AUG2019

Under the guidance of
Mr. Jayveer Nanda

CONTENTS

Chapter 1: Introduction……………………………………………………………….4

Chapter 2: Problem Statement……………………………………………………...…5

Chapter 3: Business Use Cases………………………………………………….....…6

Chapter 4: Stakeholders…………………………………………………….……..……..7

Chapter 5: Data Acquisition……………………..8

5.1 Data Source……………………………………………………………......8

5.2 Data Description ………………………………………………….……..8

Chapter 6: Data Preprocessing……………………...9

Chapter 7: Exploratory Data Analysis……………………..10

Chapter 8: Introduction to Time Series……………………..20

8.1 What is Time Series…………………………………………………….20

8.2 Decomposition ………………………………………………………….20

8.3 Stationarity ……………………………………………………………...20

Chapter 9: Introduction to ARIMA Model……………………......................................23

Chapter 10: Model Building……………………...25

10.1 Forecast Model for Inbound Passenger Traffic…………….…………25

10.2 Forecast Model for Outbound Passenger Traffic ……………….……..34

10.3 Forecast Model for Inbound Freight Traffic ……………………….…43

10.4 Forecast Model for Outbound Freight Traffic ………………….……..52

Chapter 11: Model Evaluation……………………...60

Chapter 12: Future Traffic Projections……………………..61

12.1 Total Traffic Projections………………………………………….....61

12.2 Projections of Inbound and Outbound Passenger Traffic……………..62

12.3 Projections of Inbound and Outbound Freight Traffic…………….….64

Chapter 13: Areas to Focus……………………..66

13.1 Cargo Services………………………………………………………....67

13.2 Human Resource Management…………………………………….…67

13.3 The Gulf Corridor……………………………………………………..68

13.4 Skill Enhancement…………………………………………………....69

13.5 Aviation Security……………………………………………….….…70

13.6 Manufacturing………………………………………………………..71

Chapter 14: Model Limitations…………………………..……..72

Chapter 15: Factors Influencing Future Demand………...73

Chapter 16: Enhancements in Future Forecast Models…..74

Chapter 17: Conclusion…..75

Chapter 18: References…...76

Acknowledgement

The final outcome of this project required a lot of guidance and assistance from many people and we are extremely fortunate to have got this all along while completing our project work. Only because of such guidance and mentoring we could accomplish the project and so we extremely thankful and grateful for that.

We owe our profound gratitude to our project guide **Mr. Jayveer Nanda,** who took keen interest on our project work and guided us all till the project was completed by providing all the necessary information and encouraging us continuously. The completion of this report would not have been possible without such guidance and support.

We would also like to thank the teaching staff of department of Data Science Engineering for helping us in successfully completing our project work.

Chapter 1: Introduction

The aviation industry has seen a lot of changes over the last few decades. Aviation has become an essential means of transportation in terms of leisure or business travelling and increasing the connectivity worldwide. India's aviation sector has increasingly established itself as a safe, affordable and credible alternative. India's aviation sector supports 7.5 million jobs directly and indirectly in the country and makes up 30 billion rupees of its gross domestic product, or 1.5 percent of the economy, according to the International Air Transport Association. There are seven major airlines that dominate the market, with three full-service companies including the debt-laden, state-owned Air India, and budget operators such as IndiGo and SpiceJet.

With so many airlines, there is rising competition, and as companies add more flights, it has become increasingly difficult to be profitable. The aviation sector has rapidly grown over the past years in India, putting new development constraints on the country's old and congested airports. On a global scale, passenger air travel is expected to maintain positive growth rates up to 2030, despite a number of challenges faced by the industry. Airlines around the world are struggling with high jet fuel prices and sluggish economic growth. However, these difficult economic conditions are predicted to be offset by an increase in passenger figures, which in turn is projected to translate into the improved financial performance of the aviation industry.

The industry must continue to work effectively with all the key stakeholders, including the government and policy makers to ensure that the increase in demand can be met by optimizing the operational capabilities of the Indian Airports.

Chapter 2: Problem Statement

Analyze the traffic flow at major Indian airports and forecast the demand for international air traffic of passengers and freight based on historical data to propose improvements and planning of services required to meet the demands and to solve key business problems that drive aviation management.

Chapter 3: Business Use Cases

There are a lot of challenges that the aviation industry faces. Being able find a solution to our problem will help solve many business problems in the aviation industry. Some of the key business use cases are:

- Planning of new/expanded airport facilities such as Terminal facilities/ATC towers/Support Facilities
- Financial Feasibility / Financial Planning
- Facilitate development of national and international airways systems
- Assist manufacturers in industry to anticipate levels of aircraft orders
- Improve performance of domestic airlines and increase their market share
- Estimate Revenue generation and manage operational costs of airports
- Suggest improvements in Trade Policies w.r.t the air freight demand
- Improvements in Travel and Tourism competitiveness
- Continued improvement in the Ease of Doing Business will help or a sustainable growth of aviation
- Help the government and other agencies to make decisions on investment policies

Chapter 4: Stakeholders

This project can play a very vital role in terms of its utilization by the stakeholders. Some of the key stakeholders identified are:

- **Government** : The Government of India is a major stakeholder, as the government can make decisions on investment into the aviation sector and improve the trade policies and also manage the financial budgeting
- **Airport Authority of India (AAI) :** The AAI is another important stakeholder, as the forecasts with help AAI from the planning of new/expanded facilities at the Airports and other infrastructure developments
- **Manufactures in the Aviation Industry** : With the demand of the traffic, airlines would want to acquire more aircrafts and other services, and this can help the manufactures anticipate the demand and invest into production
- **Airlines :** Airlines can schedule more flights and improve other services and gain a better growth rate

Chapter 5: Data Acquisition

5.1 Data Source

The Data used for the Analysis is collected from the official website of the Director General of Civil Aviation (DGCA). The Directorate General of Civil Aviation is the regulatory body in the field of Civil Aviation primarily responsible for regulation of air transport services to/from/within India and for enforcement of civil air regulations, air safety and airworthiness standards.

Statistical Division of the Director General of Civil Aviation is responsible for maintaining civil aviation traffic statistics. This Division collects data pertaining to Civil Aviation from various sources viz. Air India, Indian Airlines, Private Operators, Foreign Airlines and various airports managed by Airports Authority of India. The data thus collected are compiled and are then published annually in a publication entitled "India Air Transport Statistics" http://www.dgca.nic.in/reports/rep-ind.htm

5.2 Data Description

The following Data is collected from the above mentioned Source

- Citypair-wise Monthly Domestic Passenger Traffic (2015-2018)
- Airline-wise Monthly International Passenger and Freight Traffic (2014-2018)
- Country-wise Quarterly International Passenger and Freight Traffic (2014-2018)
- Citypair-wise Quarterly International Passenger and Freight traffic (2014-2018)

Attributes in the Data:

- YEAR
- MONTH
- QUARTER
- AIRLINE NAME : Name of the Airlines
- PASSENGERS TO INDIA : Number of passengers who flew to India
- PASSENGERS FROM INDIA : Number of passengers who flew out of India
- COUNTRY : Name of the country
- CITY1 : Name of the international city (International Traffic Data)
- CITY2 : Name of the Indian city (International Traffic Data)
- CARRIER TYPE : Foreign or Domestic airlines
- FREIGHT TO INDIA : Amount of cargo carried to India (in Tonnes)
- FREIGHT FROM INDIA : Amount of cargo carried from India (in Tonnes)
- CITY1 : Source city (Domestic Traffic Data)
- CITY2 : Destination city (Domestic Traffic Data)

Chapter 6: Data Preprocessing

Sample of Airline-wise monthly international traffic to and from India

	YEAR	MONTH	QUARTER	AIRLINE NAME	CARRIER TYPE	PASSENGERS TO INDIA	PASSENGERS FROM INDIA	FREIGHT TO INDIA	FREIGHT FROM INDIA
0	2015	JAN	Q1	AIR INDIA	DOMESTIC	258876.0	274220	3320.626	4186.302
1	2015	JAN	Q1	AIR INDIA EXPRESS	DOMESTIC	95581.0	116600	0.000	0.000
2	2015	JAN	Q1	INDIGO	DOMESTIC	68112.0	74212	320.000	1812.000
3	2015	JAN	Q1	JET AIRWAYS	DOMESTIC	320853.0	332116	4173.874	5383.515
4	2015	JAN	Q1	SPICEJET	DOMESTIC	37882.0	42468	0.000	115.680

- The Year and Month are in different columns, both the columns should be merged to get a date column.
- Since the traffic is recorded at the end of the month, an offset should be added the last date of every month.
- To get the total Passenger and Freight traffic, both the columns 'To India' and 'From India' are added to get a new column.

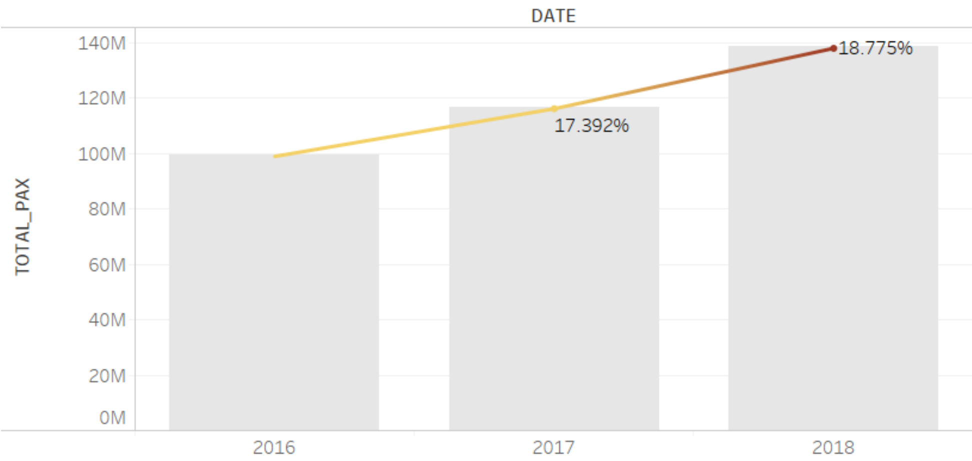
Sample of the processed data

	DATE	YEAR	QUARTER	COUNTRY_NAME	PASSENGERS_TO_INDIA	PASSENGERS_FROM_INDIA	FREIGHT_TO_INDIA	FREIGHT_FROM_INDIA	TOTAL_PASSE
0	2015-03-31	2015	Q1	AFGHANISTAN	37194	32721	96.7	676.4	
1	2015-03-31	2015	Q1	AUSTRALIA	16676	19075	89.7	423.7	
2	2015-03-31	2015	Q1	AUSTRIA	13318	15529	563.8	707.0	
3	2015-03-31	2015	Q1	BAHRAIN	114087	133527	585.7	1909.5	
4	2015-03-31	2015	Q1	BANGLADESH	76357	75673	624.1	1052.9	

- The processed dataframe has the date column with the added offset and also the total passenger and freight traffic.
- The same is applied to all the other data files to get the date and total columns.
- The data does not contain any Missing Values

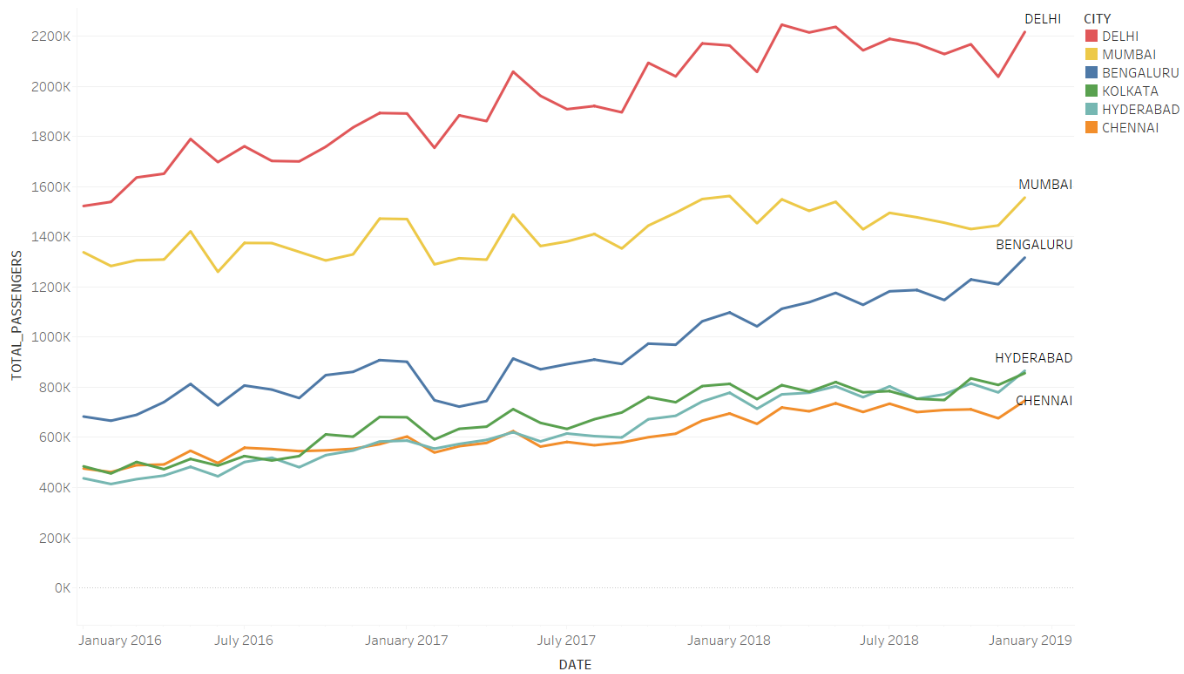
Chapter 7: Exploratory Data Analysis

Annual Growth of Domestic Passenger Traffic



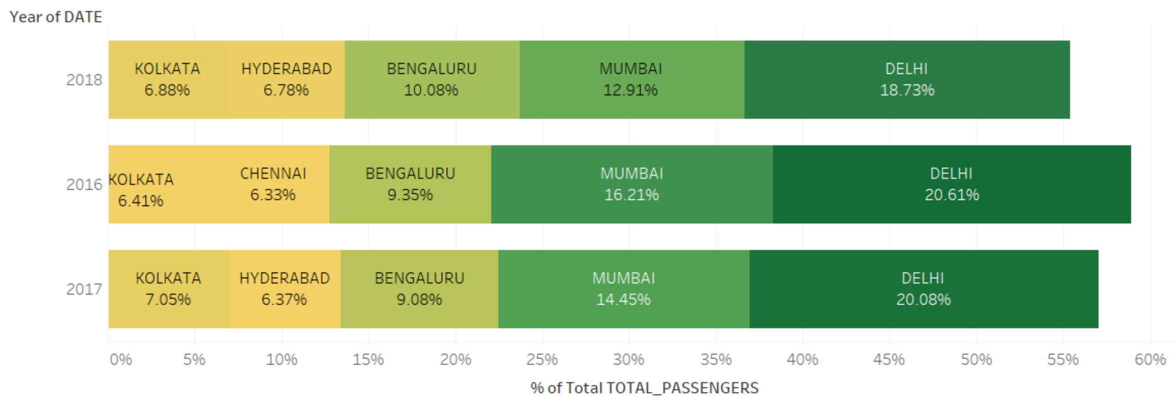
- There is a huge increase in the YoY growth of the Domestic Passenger Traffic for
- 2018 recorded one of the best growth rate of 18.7%

Domestic Passenger Traffic at Metro Cities



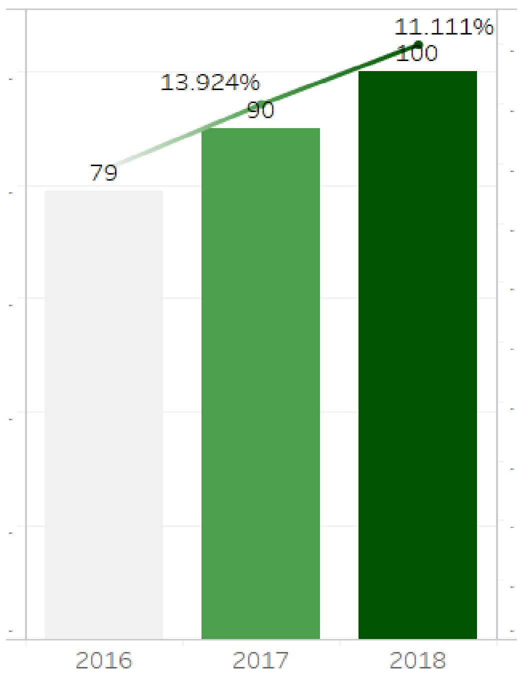
- The growth rate of Bangalore airport in the highest among all the other metro cities
- Delhi airport saw a very high growth rate till 2018, but reduced a little in 2018
- Mumbai airport seems to have a very slow growth rate
- Hyderabad, Chennai and Kolkata all seem share almost equal traffic with a decent growth rate

Market Share of top 5 Cities on Domestic Passenger Traffic



- Delhi airport has the highest share of domestic passenger traffic of about 20%, followed by the Mumbai airport with an average of 14%
- The can be understood as Delhi is the legislative capital and Mumbai is the financial capital of India. We can always anticipate a high traffic rate at these two airports

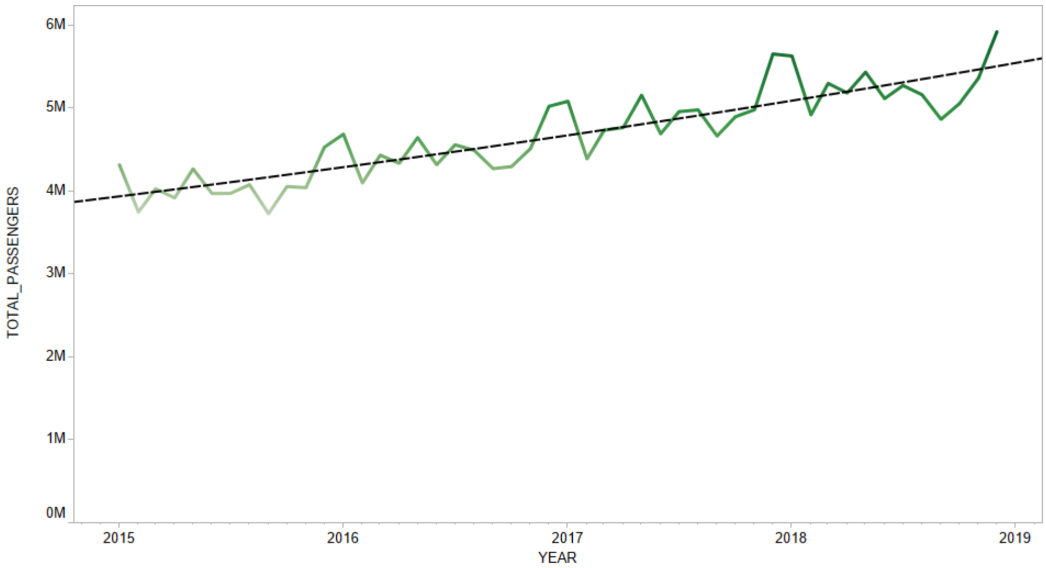
Total Airports Handling Domestic Passenger Traffic

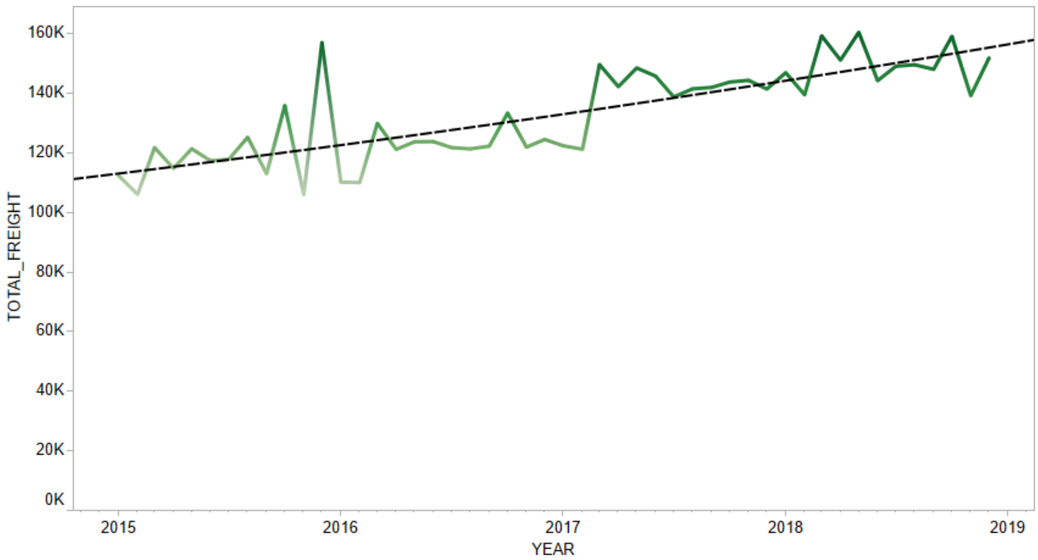


Over the last 3 years there is an overall increase of 25% in the total number of airports that are handling domestic traffic. This suggests that more number of cities are connected by the air transport routes and also suggest a significant growth in the aviation industry.

The increase in the connectivity is an example of the **UDAN** (Ude Desh ka Aam Naagrik) scheme launched by the Government of India for the development of airports and regional connectivity, letting the common citizen of the country fly and boost national economic development, job growth and air transport infrastructure development.

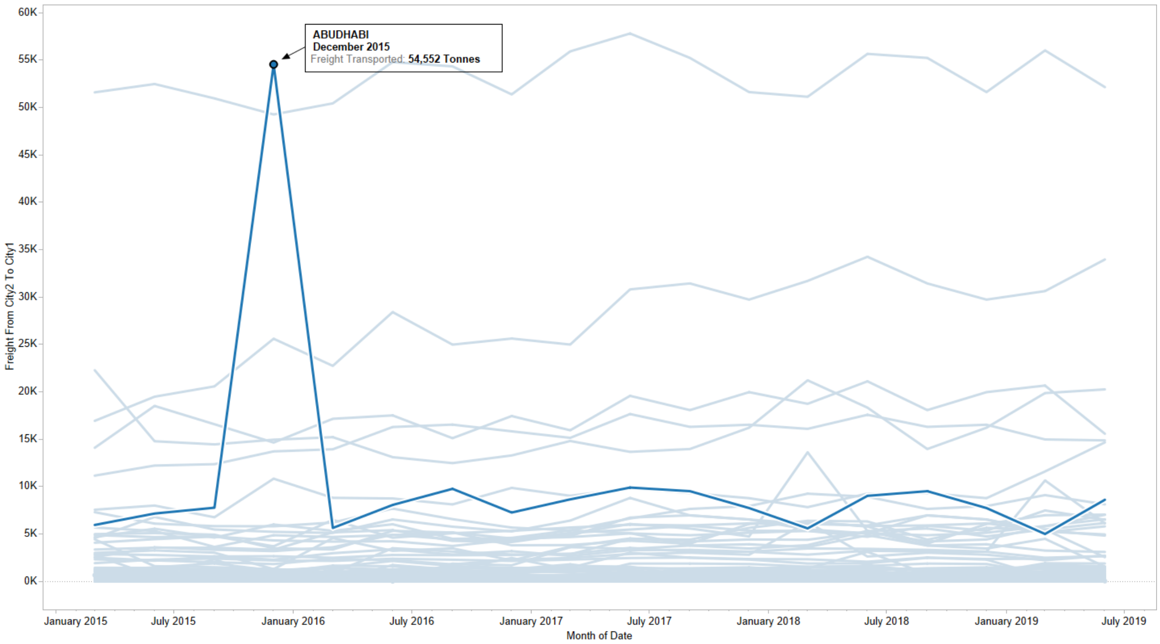
Total International Passenger and Freight (in Tonnes) Traffic





- The total passenger traffic to and from India reached to 6 Million monthly in 2018, which is the highest of all.
- There is a clear linear up trend in the increase of passenger and freight traffic to and from India over the last 4 years.
- The total Freight traffic increased from less than 120K Tonnes to more that 150K Tonnes by 2018.

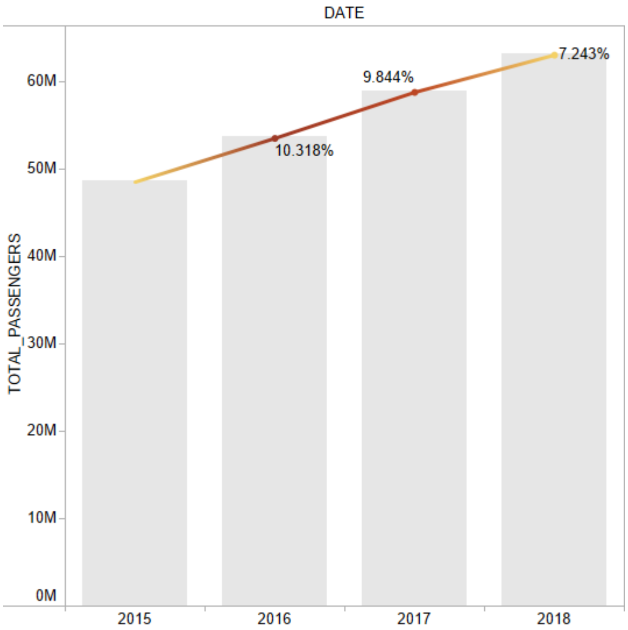
We can see a huge spike in the Freight Traffic in the month of December 2015, upon further investigation, we find the following Insight:



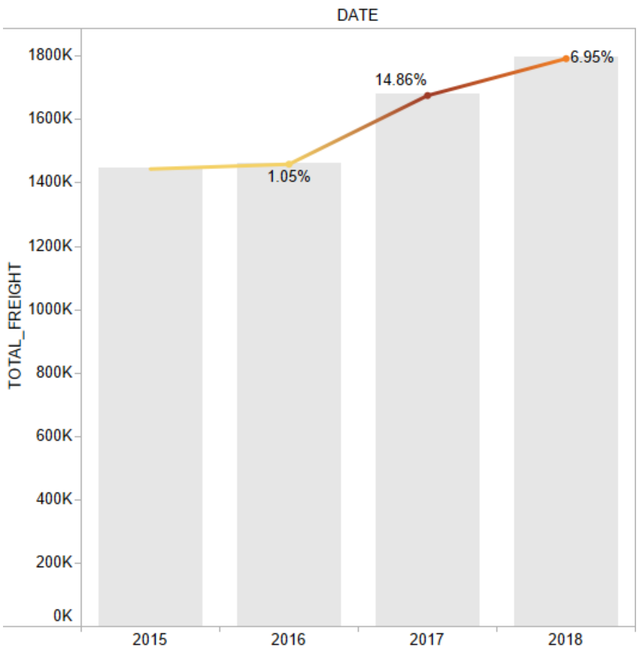
The huge spike in the graph for Freight Traffic, was due to the fact that the Louvre Museum was announced to be opened in Abu Dhabi. India had a strategic partnership and agreement with France and the Louvre Museum with the aim of establishing an active partnership in the area of exchange of competencies and expertise, particularly in the field of museology, temporary exhibitions. Hence huge amount of cargo was transported between Abu Dhabi and India.

This suggests that, establishing and maintaining a strategic relationships with the other countries helps us improve our trade. This in turn can be focused on improving the Trade in terms of the aviation sector and the air services to those countries, which would help India to growth with more pace.

Annual Growth of International Passenger and Freight Traffic

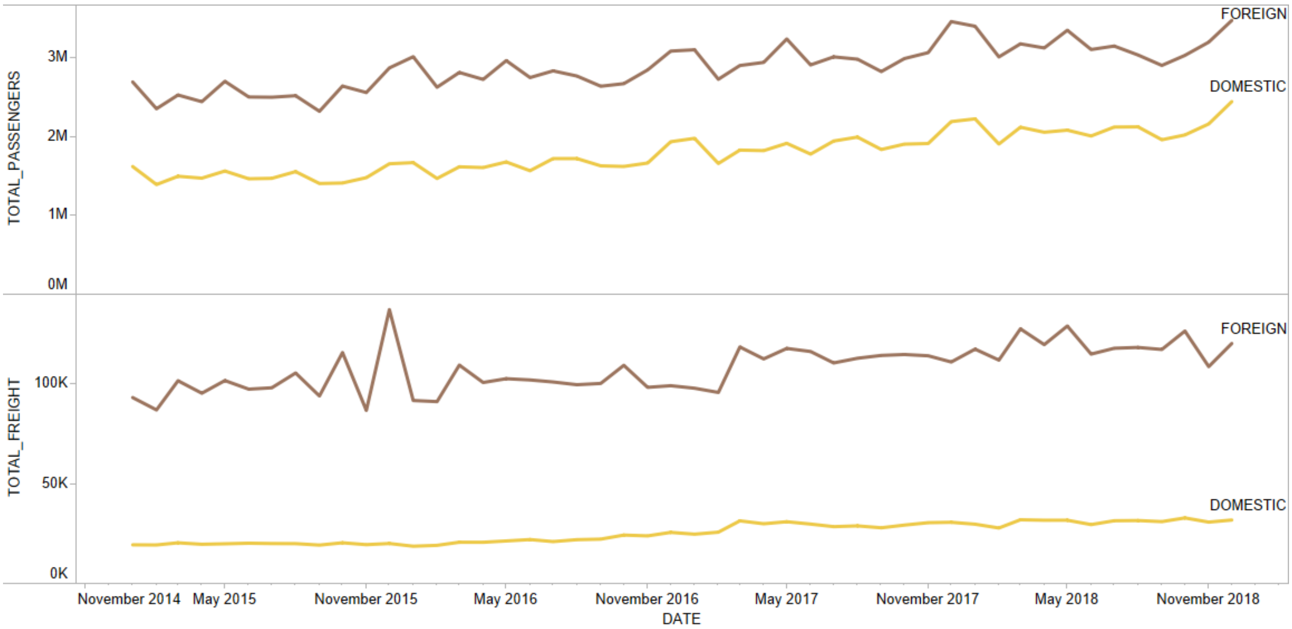


- There is a good growth rate of close to 10% YoY
- International Passenger Traffic increased close to 65 Million in 2018



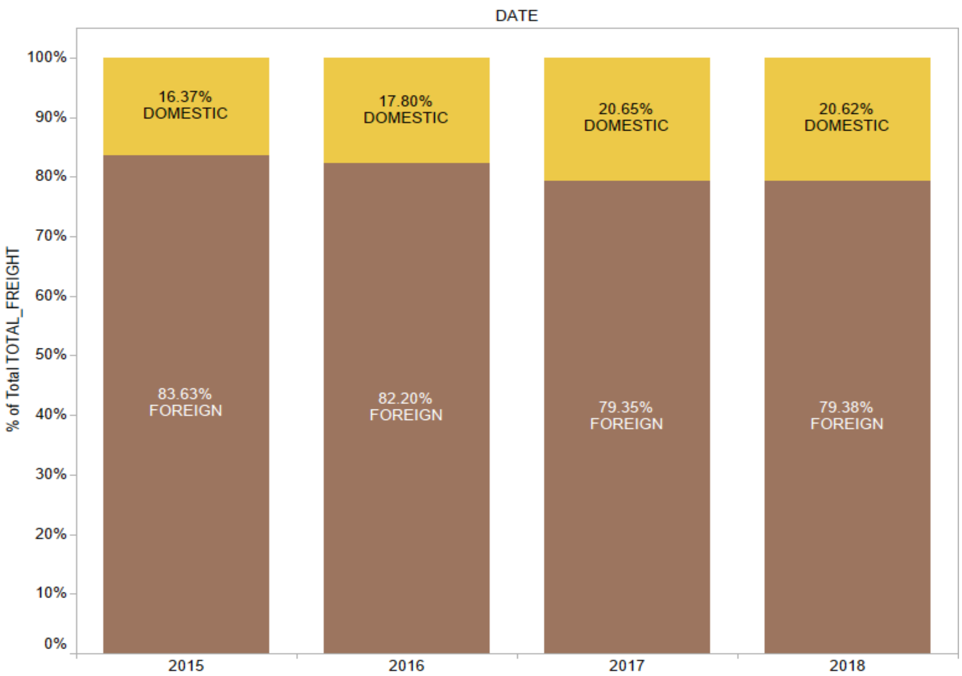
- 2017 saw the highest growth rate of more than 14% in Freight traffic
- This suggests a huge demand in the Air Cargo Traffic for India

International Traffic carried by Domestic vs. Foreign Carriers



- In terms of Passenger traffic, Foreign carriers carry almost twice the traffic as Domestic carriers.
- The growth rate of both Domestic and Foreign carriers seem to increase in the same rate in terms of Passenger traffic.
- In terms of Freight traffic, we can see a study growth in Foreign carriers whereas there is almost no growth in Domestic carriers

Market Share of Freight traffic by Carrier Type

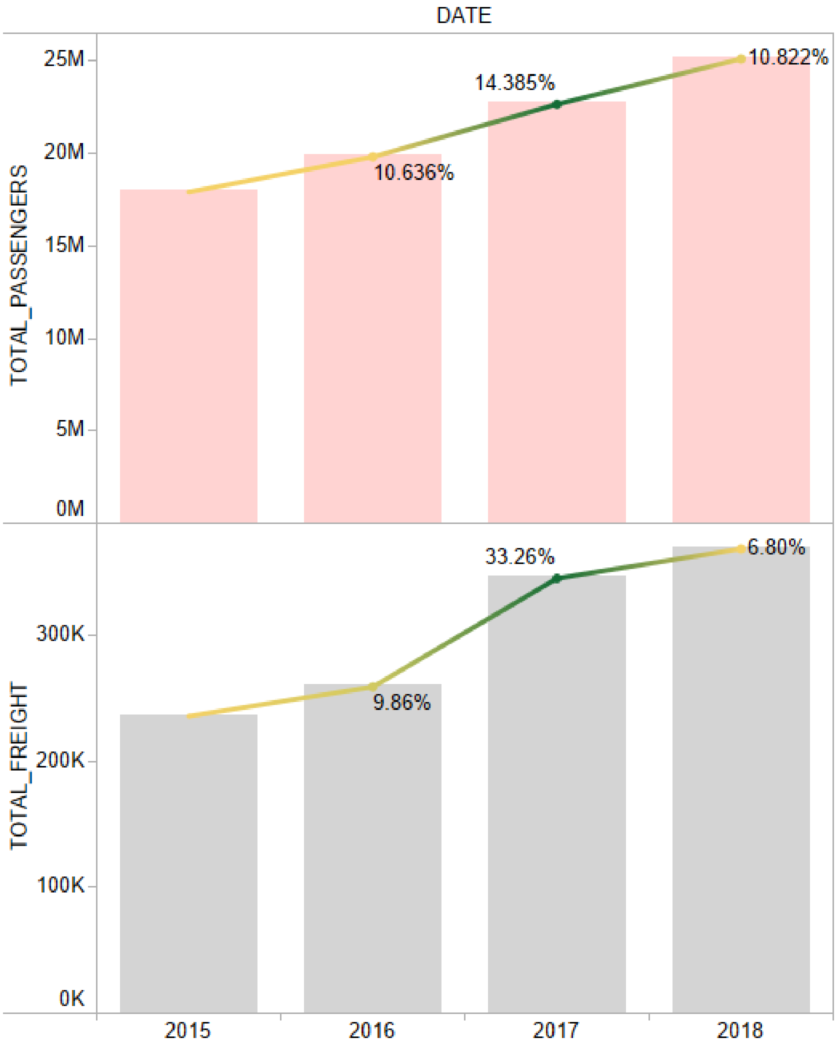


- On an average ever year only 20% of the total Freight traffic is carried by the Domestic carriers, which implies the Market Share of Domestic carries in very less in terms of Freight traffic.

- **More focus should be given to the Cargo Services of the Domestic carriers to capture more traffic and increase the market share.**

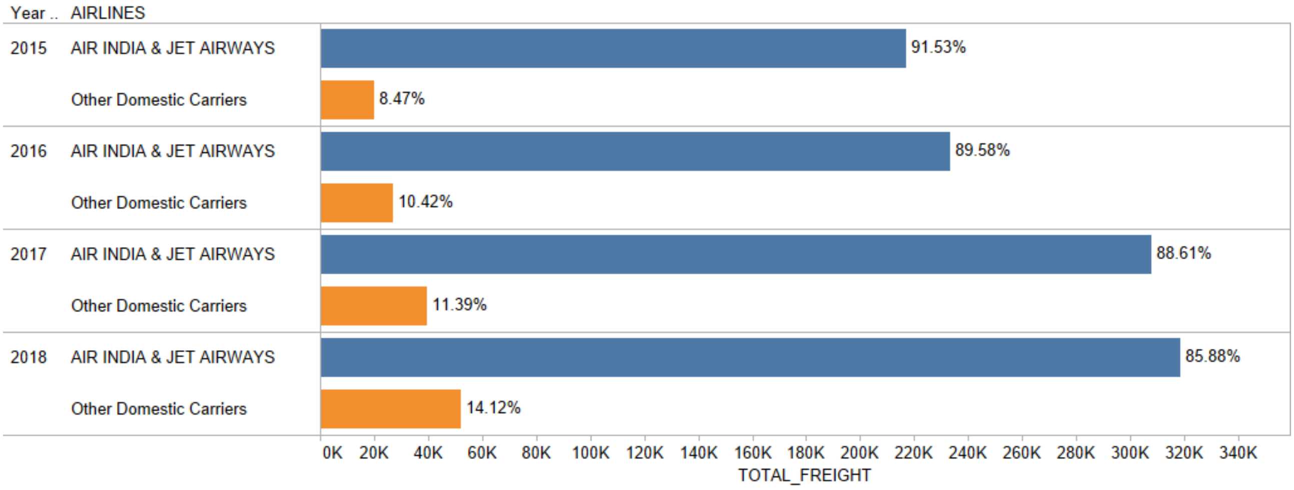
- **Increasing the market share of Domestic carriers will not only lead to growth of the carriers but also attract more trade routes and increase in demand and revenue.**

Domestic Carrier Growth (International Traffic)



- There is more than 10% increase in the Passenger traffic carried by the Domestic carriers every year.
- There is an overall increase of more than 50% in the Freight traffic carried by the Domestic carriers over the last 4 years.
- If the services of the Domestic carriers are improved, there is a huge potential of growth for the carriers and for the aviation sector in India.

Share of Freight Traffic within Domestic Carriers



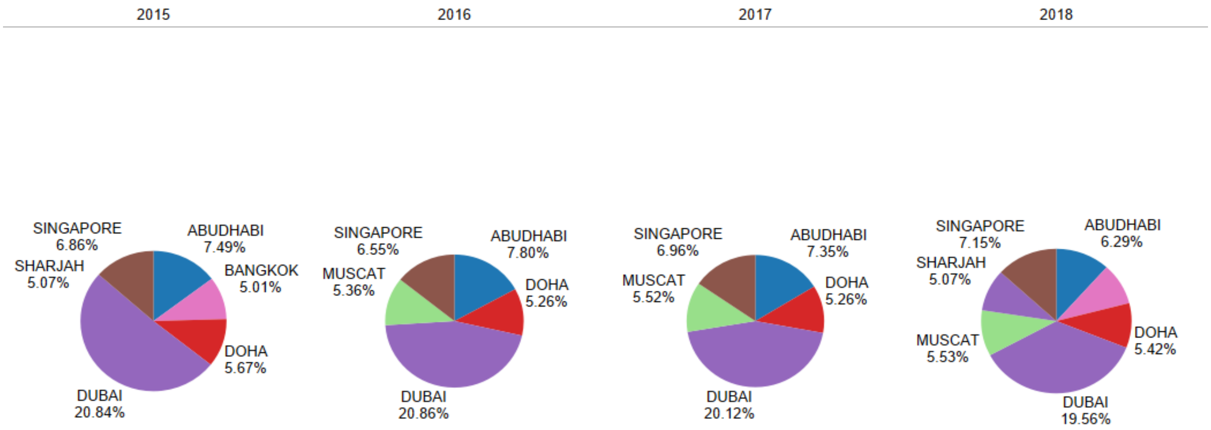
- Air India and Jet Airways together carry more than 85% of the total Freight traffic
- Other premier Domestic carriers such as Indigo and Spicejet carry only about 10% of the Freight traffic
- **With Jet Airways now being shutdown, the Cargo Services for the Domestic services have taken a huge blow in the market.**
- **Premier Domestic carriers such as Indigo and Spicejet should now invest more in their Cargo Services.**
- **Investment in the cargo services by the domestic carriers would be a big gain for the aviation's manufacturing sector as more aircrafts dedicated to the cargo services would be required.**
- **This could be a huge growth opportunity for the Indian market and the aviation sector.**

Due to the increased tensions between the USA and CHINA with the Trade War going on, we can see a huge fall in the demand of the Air Cargo Traffic. This could be the right time for India to step up our International Cargo Services and Operations and capture most of the traffic. It can be anticipated from the analysis that, improving our Trade policies will increase the demand for the Air Cargo Traffic for India and a high opportunity for the Domestic Indian Carriers.

Inbound and Outbound International Passenger Traffic

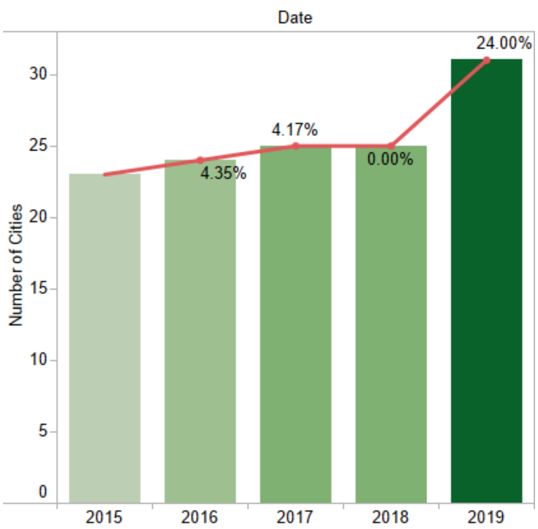
- There seems to be a pattern in the Inbound passenger traffic forming a diamond shape, every year in 4th Quarter the Inbound Passenger Traffic is more than the Outbound Passenger Traffic.
- This can be suggested as more people fly to India for the Holidays as 4th Quarter is from (October to December) and people fly out of India during in month of January
- This information can we a lot of help to the Airlines, as they can anticipate the same pattern every year and schedule more inbound flights during Q4 and more outbound flights during January.
- Scheduling the flights properly with help the airlines carry more traffic and also build a proper pricing strategy in increase their revenue.
- Since these patterns seem to occur every year, we can develop new strategies for the Tourism sector to attract more tourists and other business strategies within the airports and also outside.

Share of Passenger Traffic by International Cities



- Dubai has the highest passenger traffic to and from India of about 20% every year
- Looking at the pie charts, we can see that most of the cities with high share of passenger traffic are in the Arab Countries.
- This suggests that most the traffic flows through the Middle Eastern countries and the airlines can use this information to plan routes though these cities connecting other cities in other countries
- Since these routes have more traffic, we can develop new airways or build strategies to maintain these airways for efficient and safe routing of the flights.

Connectivity to Indian Cities



There is a huge increase in the number of cities connected to International air routes. More than 30 cities are now handle international traffic. This suggests there is a huge development the in aviation sector in India and also increasing in regional connectivity across all the Indian States.

Chapter 8: Introduction to Time Series

8.1 What is Time Series?

Time series is a sequence of observations recorded at regular time intervals. Depending on the frequency of observations, a time series may typically be hourly, daily, weekly, monthly, quarterly and annual. Sometimes, you might have seconds and minute-wise time series as well, like, number of clicks and user visits every minute etc.

Why analyse a time series?

Because it is the preparatory step before you develop a forecast of the series. Besides, time series forecasting has enormous commercial significance because stuff that is important to a business like demand and sales, number of visitors to a website, stock price etc. are essentially time series data.

Time series analysis involves understanding various aspects about the inherent nature of the series so that you are better informed to create meaningful and accurate forecasts.

8.2 Decomposition

Time series data can exhibit a variety of patterns, and it is often helpful to split a time series into several components such as trend, seasonality, cyclic, irregularity(or residue), each representing an underlying pattern category which can be done by decomposition.

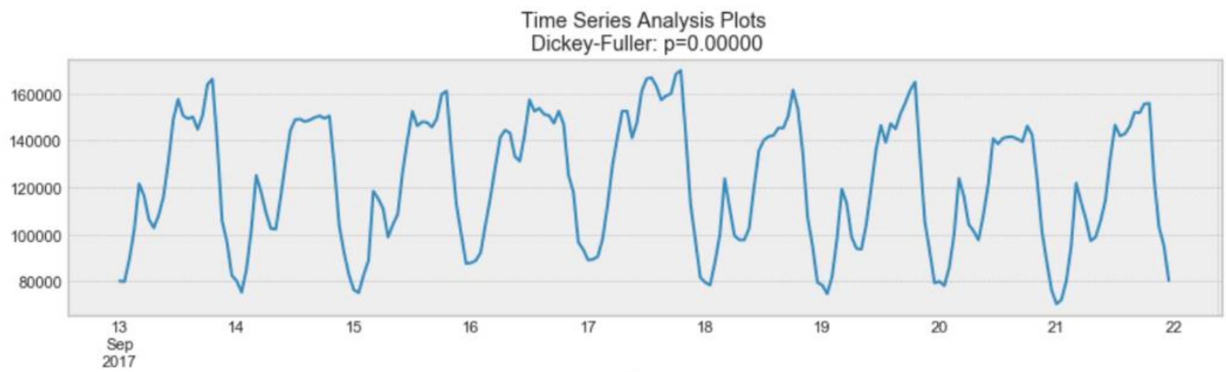
Components:

Considering the effects of these four components, two different types of models are generally used for a time series:

- Additive Model $Y(t) = T(t) + S(t) + C(t) + I(t)$
  Assumption: These four components are independent of each other.
- Multiplicative Model $Y(t) = T(t) \times S(t) \times C(t) \times I(t)$
  Assumption: These four components of a time series are not necessarily independent and they can affect one another.

8.3 Stationarity

A time series is said to be stationary if its statistical properties do not change over time. In other words, it has constant mean and variance, and covariance is independent of time.



Example of a stationary process

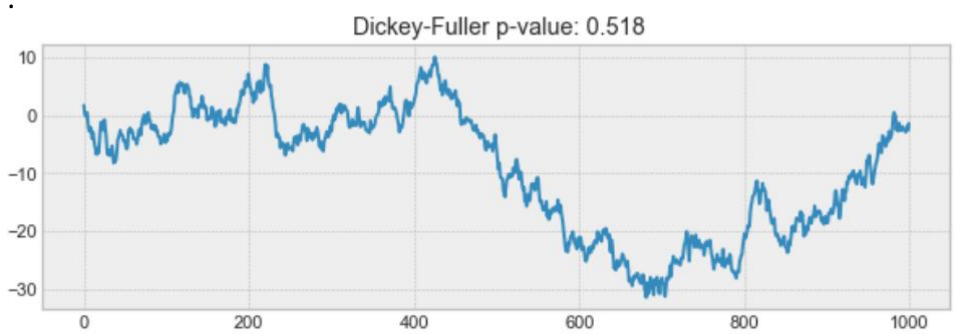
Testing Stationarity

Augmented Dickey-Fuller is the statistical test to determine if a time series is stationary or not
by checking the presence of unit root in the series. The null and alternate hypothesis of this test is:

Null Hypothesis: The series has a unit root meaning it is non-stationary.
Alternate Hypothesis: The series has no unit root meaning it is stationary.

We interpret this result using the p-value from the test.
p-value > 0.05: Fail to reject the null hypothesis (H0), the data has a unit root and is non-stationary.
p-value <= 0.05: Reject the null hypothesis (H0), the data does not have a unit root and is stationary.

.



Example of a non-stationary process

Making Time Series Stationary

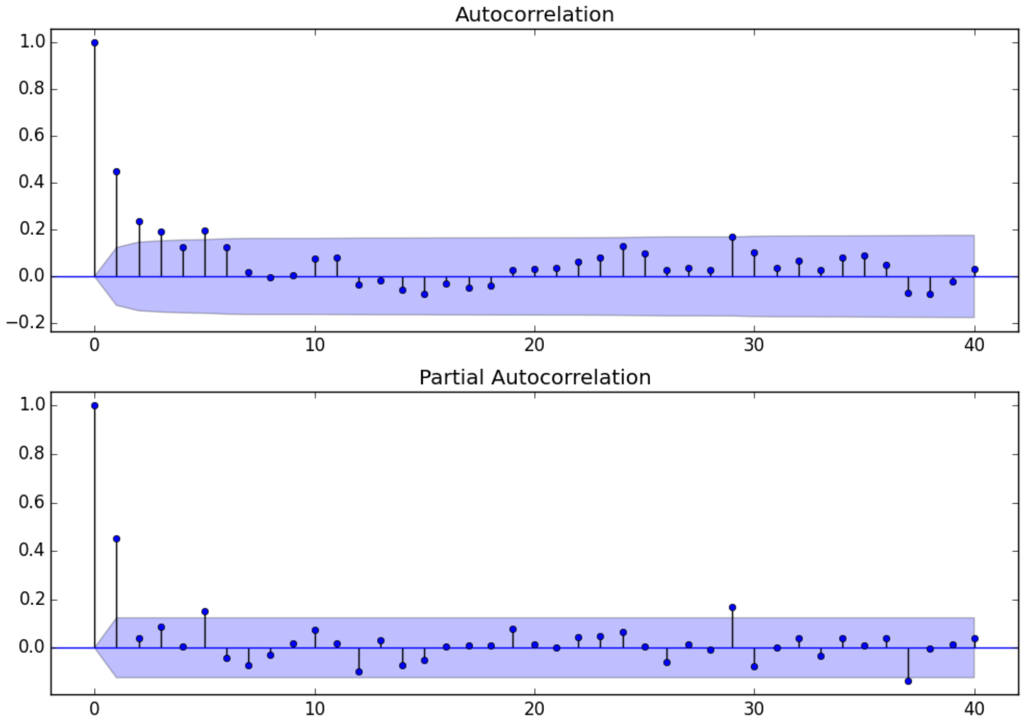
- **Log Transformation:** Exponential distribution can be made linear by taking the logarithm of the values.
- **Log Minus Weighted Average:** Log values subtracted by varying degrees of importance of the numbers in a data set.
- **Log Minus Moving Average:** Log values subtracted by an average of raw observations in the original time series
- **Log Differencing:** Computing the differences between the consecutive logarithm of the observations.
- **Differencing:** Compute the differences between consecutive observations.

Autocorrelation(ACF)

Autocorrelation is when a time series is linearly related to a lagged version of itself. It is is simply the correlation of a series with its own lags. If a series is significantly autocorrelated, that means, the previous values of the series (lags) may be helpful in predicting the current value.

Partial autocorrelation function(PACF):

The relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed.



Chapter 9: Introduction to ARIMA Model

Auto-Regressive Integrated Moving Average (ARIMA)

Autoregression(AR): A model that uses the dependent relationship between an observation and some number of lagged observations.

Integrated(I): The use of differencing of raw observations in order to make the time series stationary.

Moving Average(MA): A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.
AR and MA are two widely used linear models that work on stationary time series, and I is a preprocessing procedure to “stationarize” time series if needed.

Parameters of ARIMA model:

- p - The number of lag observations included in the model, also called the lag order.
- d - The number of times that the raw observations are differenced, also called the degree of differencing.
- q - The size of the moving average window, also called the order of moving average.

A problem with ARIMA is that it does not support seasonal data. That is a time series with a repeating cycle. ARIMA expects data that is either not seasonal or has the seasonal component removed, e.g. seasonally adjusted via methods such as seasonal differencing. An alternative is to use SARIMA.

Seasonal Autoregressive Integrated Moving Average (SARIMA)

SARIMA is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.

Configuring SARIMA

Selecting hyperparameters for both the trend and seasonal elements of the series.

Trend Elements:

There are three trend elements that require configuration. They are the same as the ARIMA model; specifically:

- **p**: Trend autoregression order.
- **d**: Trend difference order.
- **q**: Trend moving average order.

Seasonal Elements:
There are four seasonal elements that are not part of ARIMA that must be configured:

- **P**: Seasonal autoregressive order.
- **D**: Seasonal difference order.
- **Q**: Seasonal moving average order.
- **m**: The number of time steps for a single seasonal period.

SARIMA model can be specified as: **SARIMA(p,d,q)(P,D,Q,m)**

Chapter 10: Model Building

10.1 Forecast Model for Inbound Passenger Traffic

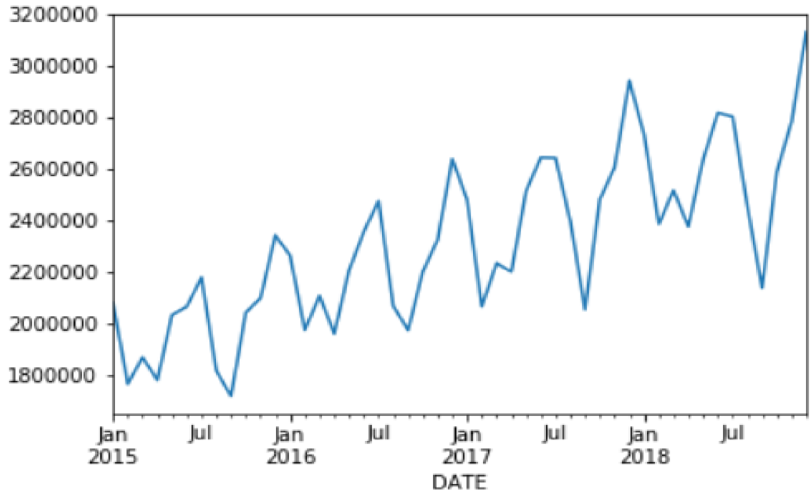
Import Data

A quick view of sample data

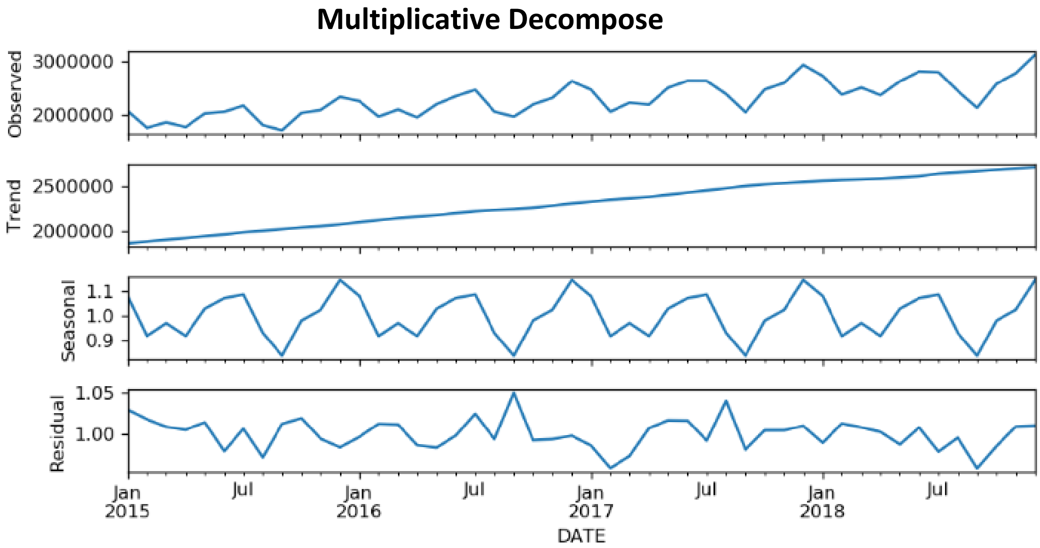
	YEAR	MONTH	QUARTER	AIRLINE NAME	CARRIER TYPE	PASSENGERS TO INDIA	PASSENGERS FROM INDIA	FREIGHT TO INDIA	FREIGHT FROM INDIA
0	2015	JAN	Q1	AIR INDIA	DOMESTIC	258876.0	274220	3320.626	4186.302
1	2015	JAN	Q1	AIR INDIA EXPRESS	DOMESTIC	95581.0	116600	0.000	0.000
2	2015	JAN	Q1	INDIGO	DOMESTIC	68112.0	74212	320.000	1812.000
3	2015	JAN	Q1	JET AIRWAYS	DOMESTIC	320853.0	332116	4173.874	5383.515
4	2015	JAN	Q1	SPICEJET	DOMESTIC	37882.0	42468	0.000	115.680

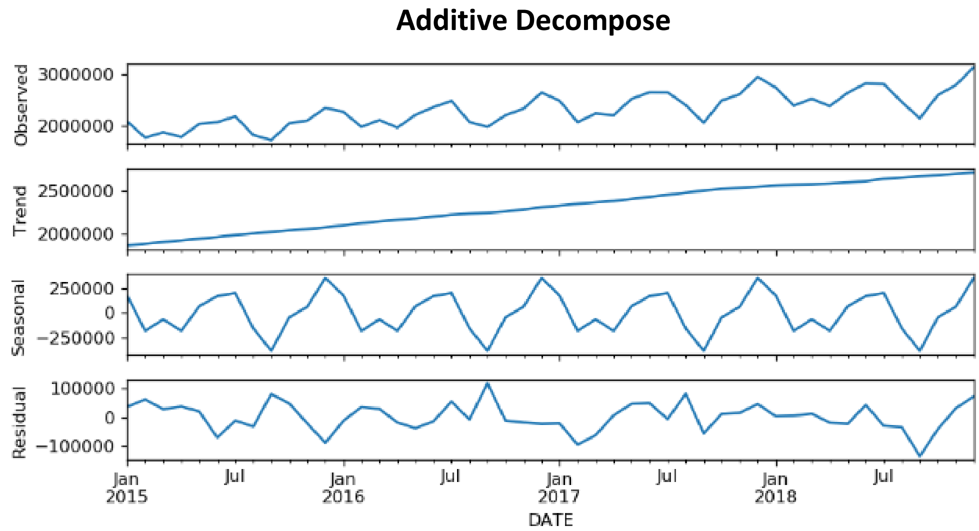
Original Series

In the beginning, it is worth looking at the general number of passengers (PAX) over time. The below graph shows PAX numbers over the years.



Decomposition



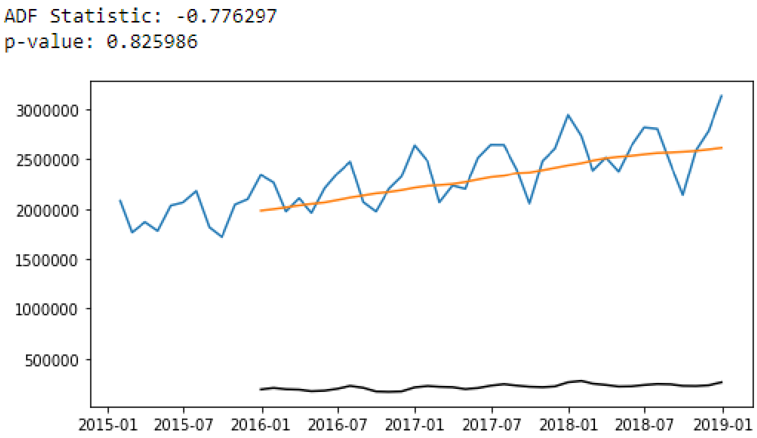


The model seems to have a 12-month seasonal trend which makes sense as usually every year a holiday period is more busy at airports.

Test for Stationarity

Let’s plot the rolling mean and rolling standard deviation having window size of 12-months over original data.

Original Series

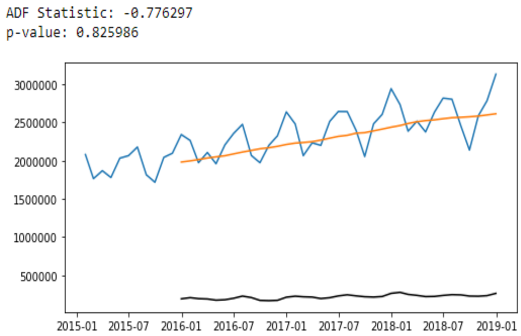


- Blue line represents our original data
- Orange line represents rolling mean and Black line represents rolling standard deviation.
- We can see that the mean and standard deviation are not constant over time. Hence the series is not stationary.
- Here p-value>0.05, so we reject the Null Hypothesis and conclude that the series is non-stationary.

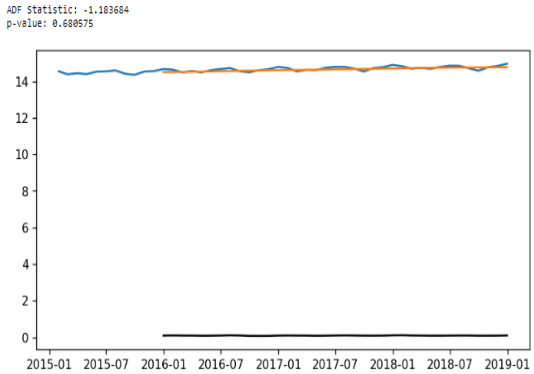
Making Time Series Stationary

In order to make time-series stationary, transform this time series by taking the 1-month difference.

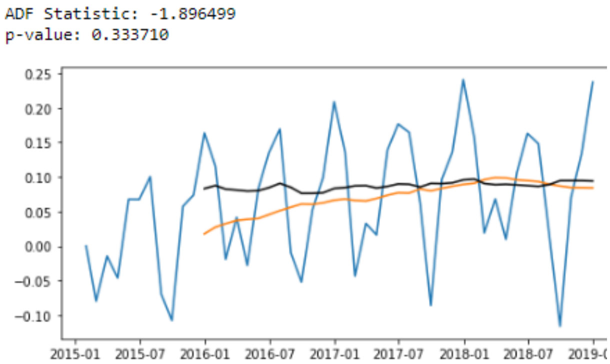
Original Series



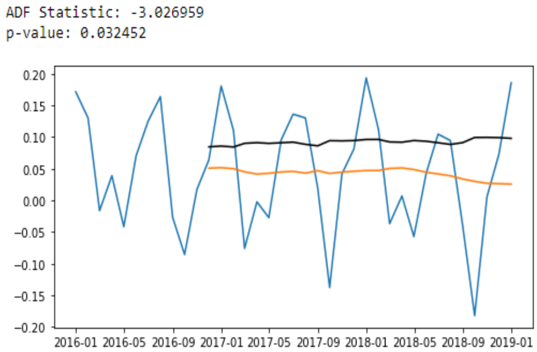
Log Transformation



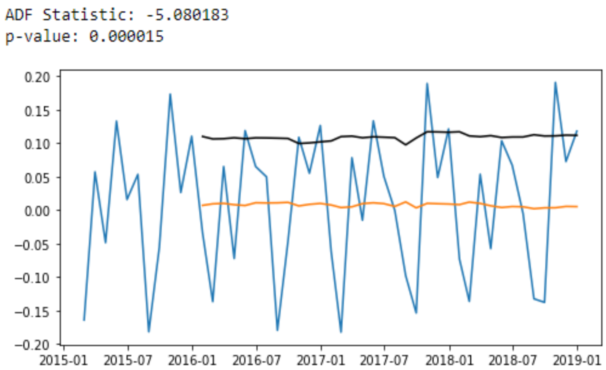
Log Minus Weighted Average



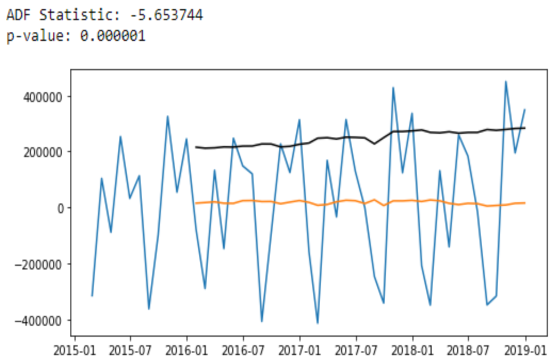
Log Minus Moving Average



LOG Differencing



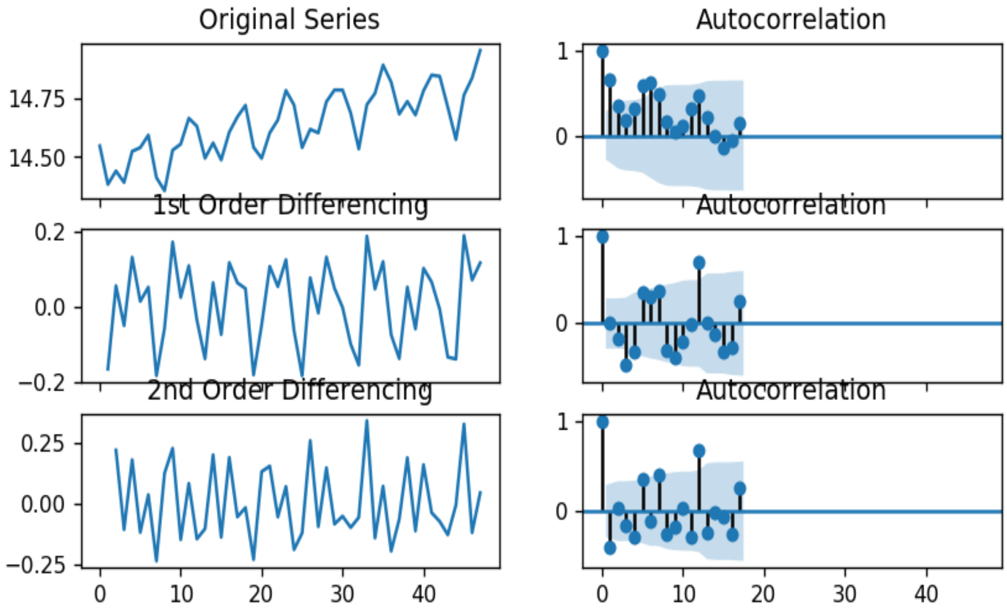
Differencing



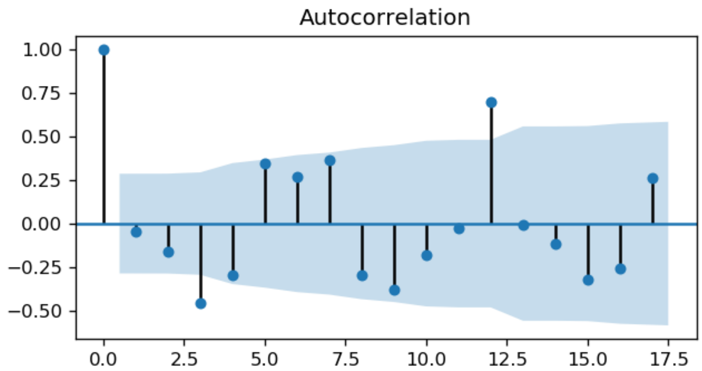
From the above graphs, we see that Log Differencing and Differencing methods have relatively stationary data having a p-value < 0.05. Hence we have transformed the time series data and are now stationary.

Finding Model Parameters (p,d,q)

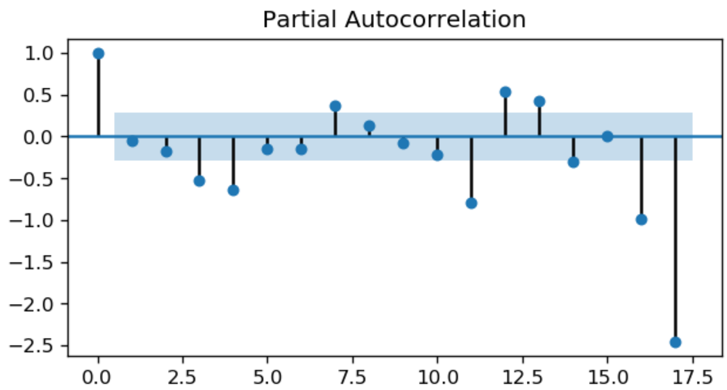
d- VALUE(N-order Differencing)



q - VALUE (AUTOCORRELATION)



p - VALUE (PARTIAL AUTOCORRELATION)



ARIMA

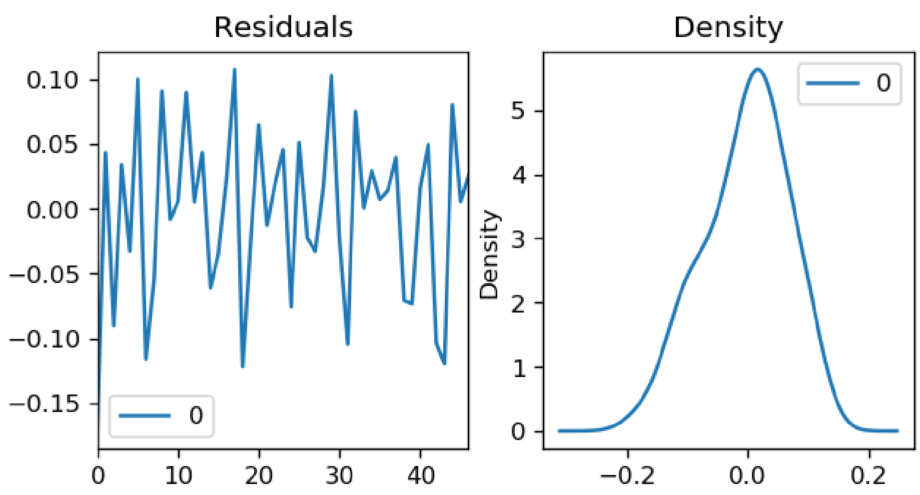
From ACF and PACF plots, we get the order as (4, 1, 0) for the ARIMA model.

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                   47
Model:                 ARIMA(4, 1, 0)   Log Likelihood                  61.467
Method:                       css-mle   S.D. of innovations              0.063
Date:                Wed, 15 Jan 2020   AIC                           -110.934
Time:                        11:53:16   BIC                            -99.834
Sample:                             1   HQIC                          -106.757

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0073      0.003      2.545      0.015       0.002       0.013
ar.L1.D.y     -0.5534      0.104     -5.329      0.000      -0.757      -0.350
ar.L2.D.y     -0.4079      0.094     -4.337      0.000      -0.592      -0.224
ar.L3.D.y     -0.6671      0.088     -7.592      0.000      -0.839      -0.495
ar.L4.D.y     -0.7068      0.103     -6.881      0.000      -0.908      -0.506
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            0.4776           -0.9200j            1.0366           -0.1738
AR.2            0.4776           +0.9200j            1.0366            0.1738
AR.3           -0.9495           -0.6442j            1.1474           -0.4051
AR.4           -0.9495           +0.6442j            1.1474            0.4051
-----------------------------------------------------------------------------
```

RMSE = 245218

Residual Diagnosis



There is a pattern in residuals and variance is not constant. The density plot is slightly left-skewed.

AUTOARIMA

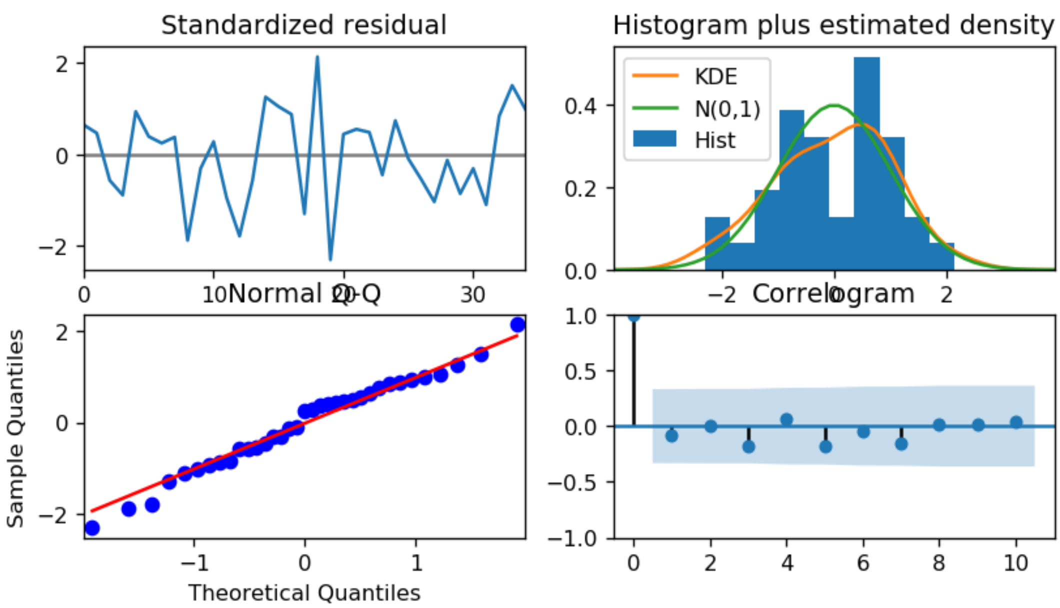
AUTOARIMA returns the best ARIMA model after trying all the possible parameters (within the constraints provided) and returns the model with the lowest AIC or BIC.

Dep. Variable:	y	No. Observations:	48
Model:	SARIMAX(0, 1, 1)x(1, 1, 0, 12)	Log Likelihood	73.826
Date:	Wed, 15 Jan 2020	AIC	-137.652
Time:	13:38:14	BIC	-129.876
Sample:	0	HQIC	-134.968
	- 48		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0127	0.011	1.132	0.258	-0.009	0.035
drift	-0.0004	0.000	-1.279	0.201	-0.001	0.000
ma.L1	-0.6294	0.162	-3.889	0.000	-0.947	-0.312
ar.S.L12	-0.4387	0.195	-2.245	0.025	-0.822	-0.056
sigma2	0.0008	0.000	3.177	0.001	0.000	0.001

Ljung-Box (Q):	nan	Jarque-Bera (JB):	0.61
Prob(Q):	nan	Prob(JB):	0.74
Heteroskedasticity (H):	1.08	Skew:	-0.25
Prob(H) (two-sided):	0.90	Kurtosis:	2.60

Residual Diagnosis



- The mean of residuals is zero except for a pattern that cannot be captured.
- The residuals are normally distributed.
- The autocorrelation in the correlogram is within the significance level.
- The residuals in QQ plot is roughly a straight line which means that it follows a normal distribution.

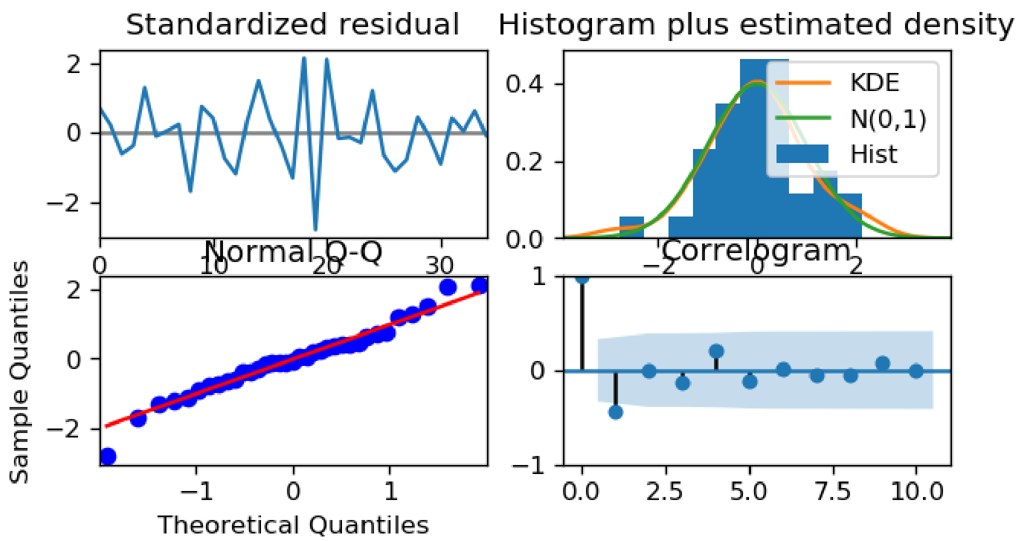
SARIMA

From AUTOARIMA, the order for SARIMA is (0,1,1)x(1,1,0,12)

```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                   48
Model:             SARIMAX(0, 1, 1)x(1, 1, 0, 12)   Log Likelihood                  73.201
Date:                            Wed, 15 Jan 2020   AIC                           -140.401
Time:                                    15:38:31   BIC                           -135.735
Sample:                                         0   HQIC                          -138.790
                                             - 48
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6289      0.140     -4.478      0.000      -0.904      -0.354
ar.S.L12      -0.4175      0.171     -2.435      0.015      -0.754      -0.082
sigma2         0.0008      0.000      3.354      0.001       0.000       0.001
===================================================================================
Ljung-Box (Q):                        nan   Jarque-Bera (JB):                 0.71
Prob(Q):                              nan   Prob(JB):                         0.70
Heteroskedasticity (H):              1.18   Skew:                            -0.18
Prob(H) (two-sided):                 0.78   Kurtosis:                         2.39
===================================================================================
```

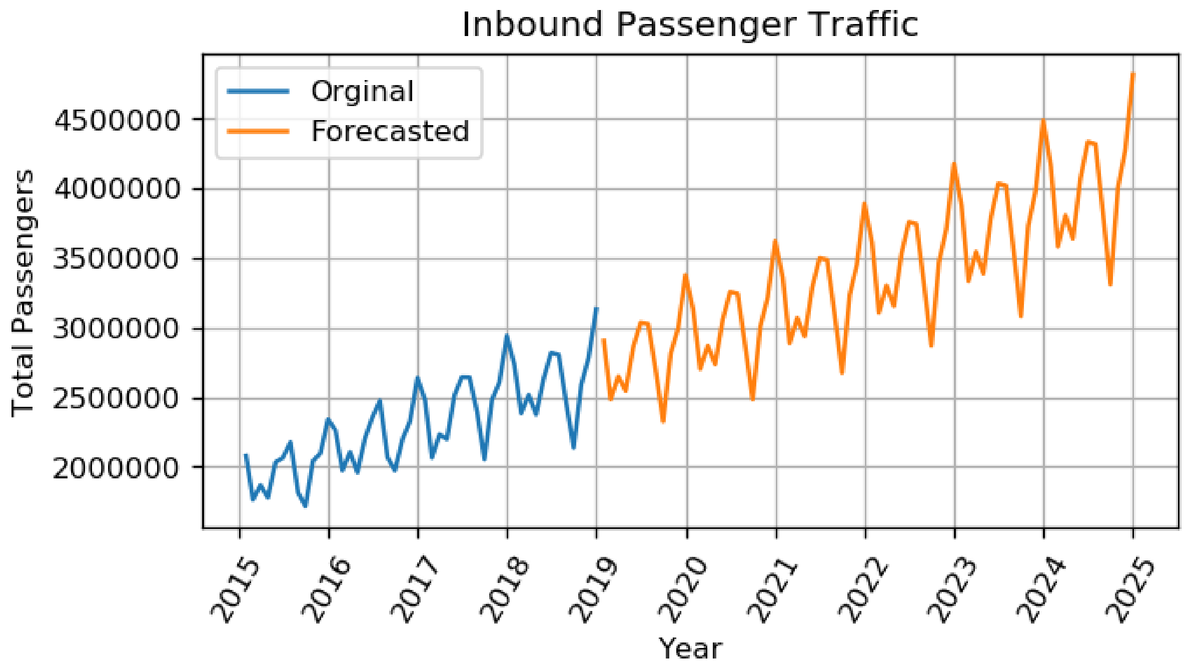
RMSE = 17362

Residual Diagnosis



- The mean of residuals is zero except for a pattern that cannot be captured.
- The residuals are normally distributed.
- The autocorrelation in the correlogram is within the significance level.
- The residuals in QQ plot is roughly a straight line which means that it follows a normal distribution.

Forecast



From the above graph, it can be observed that by 2025 the outbound passenger freight will reach up to 5 Million.

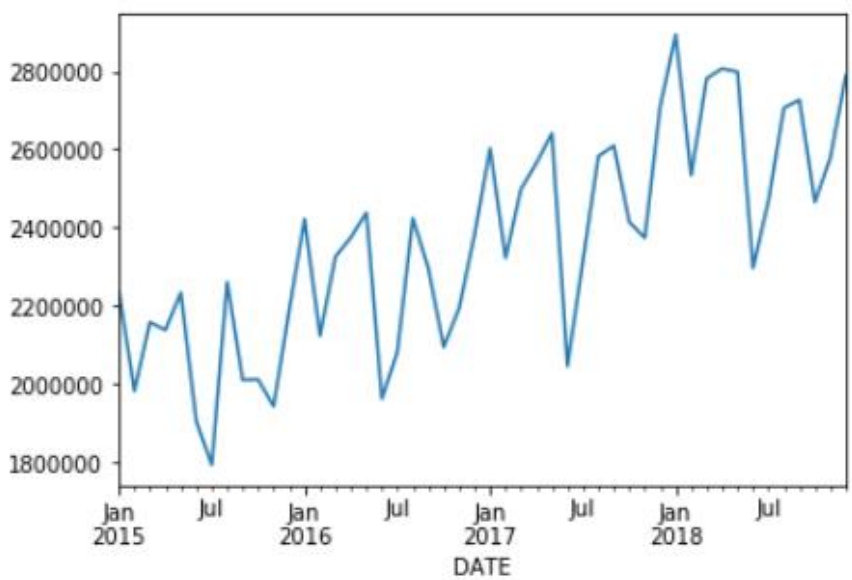
10.2 Forecast Model for Outbound Passenger Traffic

Importing Data

Here is a look at the sample data

	YEAR	MONTH	QUARTER	AIRLINE NAME	CARRIER TYPE	PASSENGERS TO INDIA	PASSENGERS FROM INDIA	FREIGHT TO INDIA	FREIGHT FROM INDIA
0	2015	JAN	Q1	AIR INDIA	DOMESTIC	258876.0	274220	3320.626	4186.302
1	2015	JAN	Q1	AIR INDIA EXPRESS	DOMESTIC	95581.0	116600	0.000	0.000
2	2015	JAN	Q1	INDIGO	DOMESTIC	68112.0	74212	320.000	1812.000
3	2015	JAN	Q1	JET AIRWAYS	DOMESTIC	320853.0	332116	4173.874	5383.515
4	2015	JAN	Q1	SPICEJET	DOMESTIC	37882.0	42468	0.000	115.680

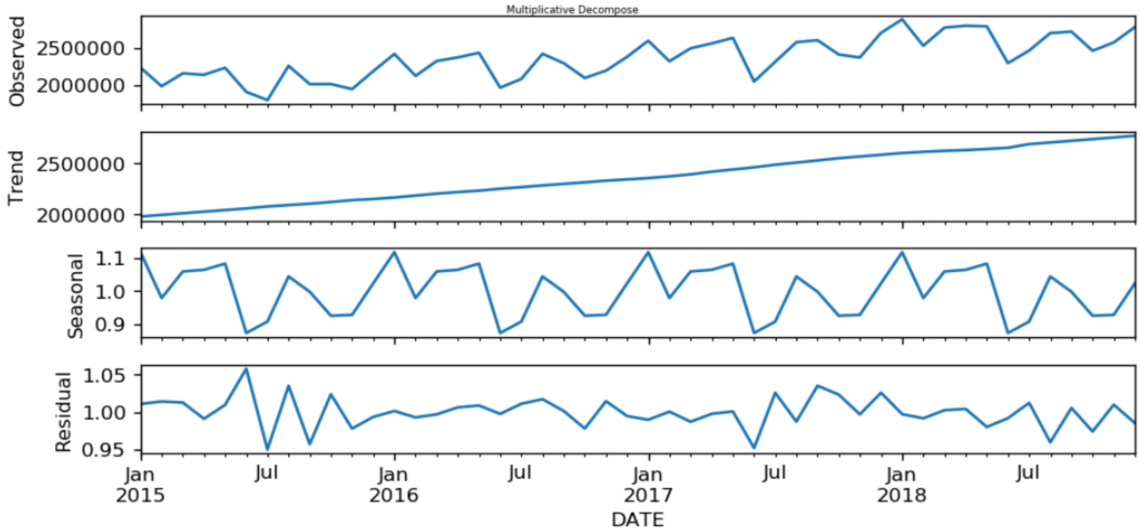
Original Series



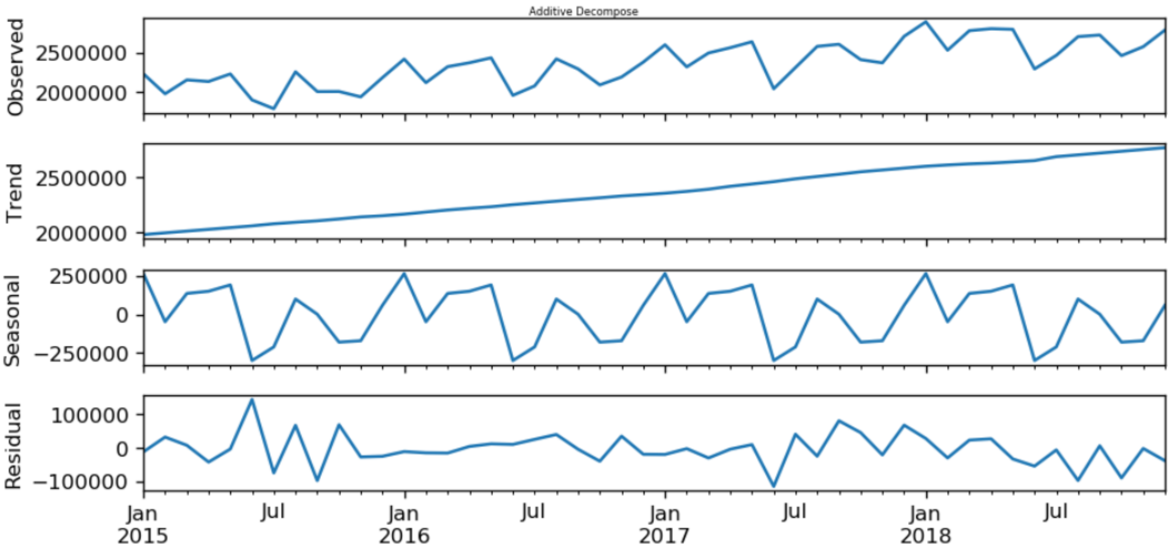
This plot gives you the number of passengers traveling from India to other countries in a month over the years.

Decomposition

Multiplicative Decompose



Additive Decompose

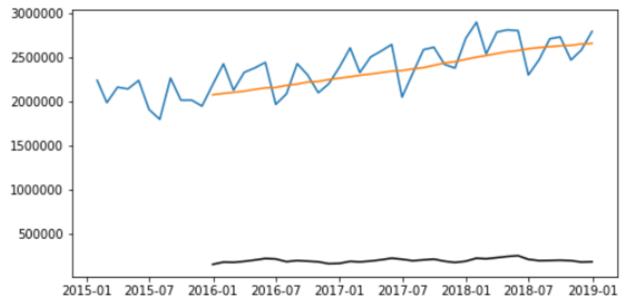


The model seems to have a 12-month seasonal trend which makes sense as usually every year a holiday period is more busy at airports.

Test for Stationarity

Let’s plot the rolling mean and rolling standard deviation having window size of 12-months over original data.

```
ADF Statistic: -0.776901
p-value: 0.825813
```

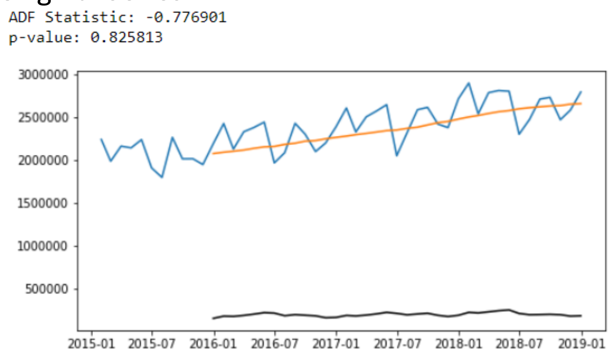


- Blue line represents our original data
- Orange line represents rolling mean and Black line represents rolling standard deviation.
- We can see that the mean and standard deviation are not constant over time. Hence the series is not stationary.
- Here p-value>0.05, so we reject the Null Hypothesis and conclude that the series is non-stationary.

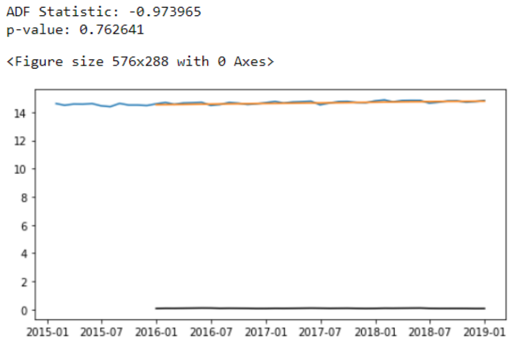
Making Time Series Stationary

In order to make time-series stationary, transform this time series by taking a month difference.

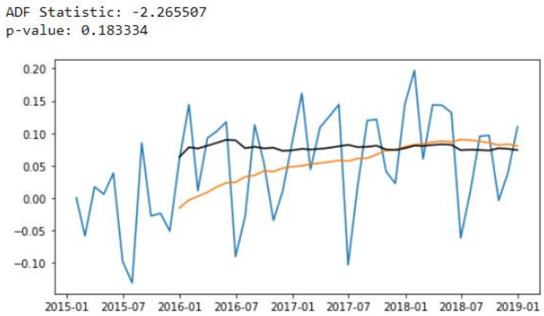
Original Series



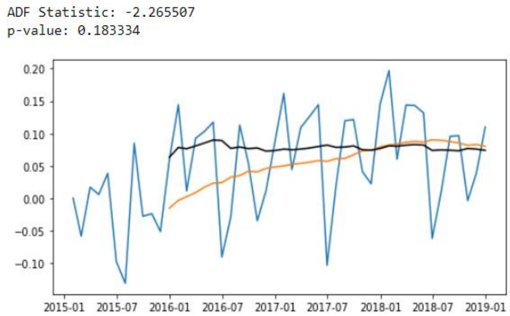
Log Transformation



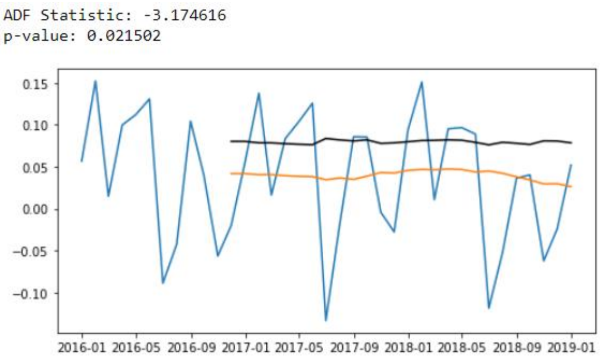
Log Minus Weighted average



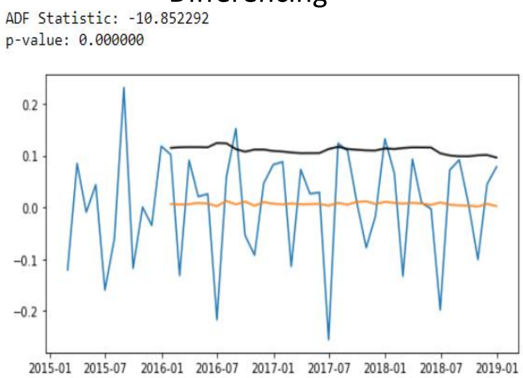
Log Minus Moving Average



Log Differencing



Differencing

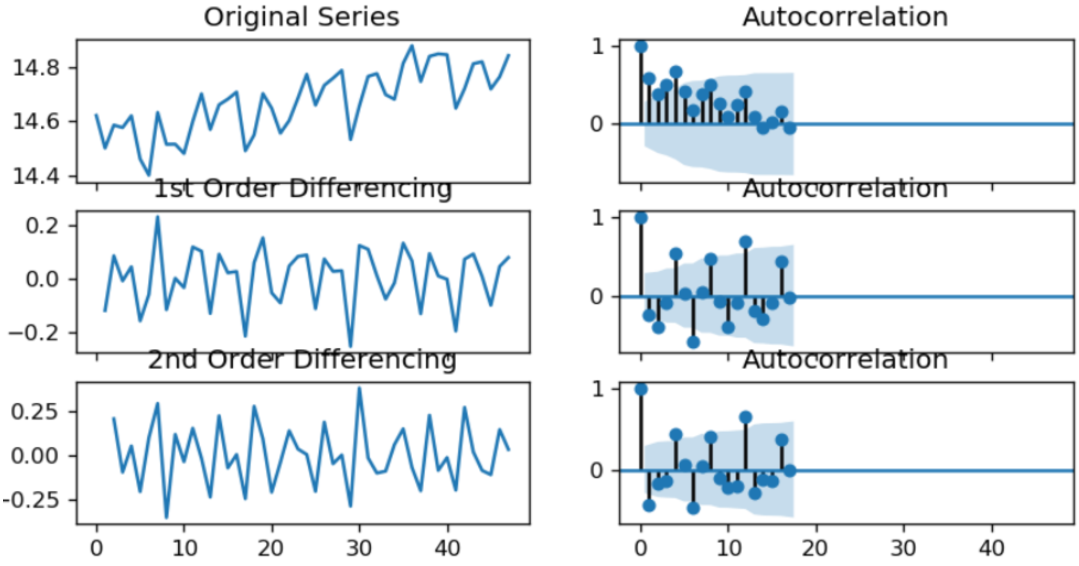


From the above graphs, we see that Log Differencing and Differencing methods have relatively stationary data having a p-value < 0.05.

Hence we have transformed the time series data and are now stationary.

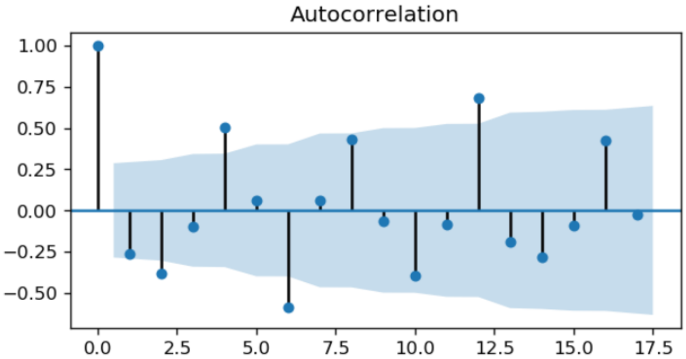
Finding Model Parameters(p,d,q)

d- VALUE(N-order Differencing)

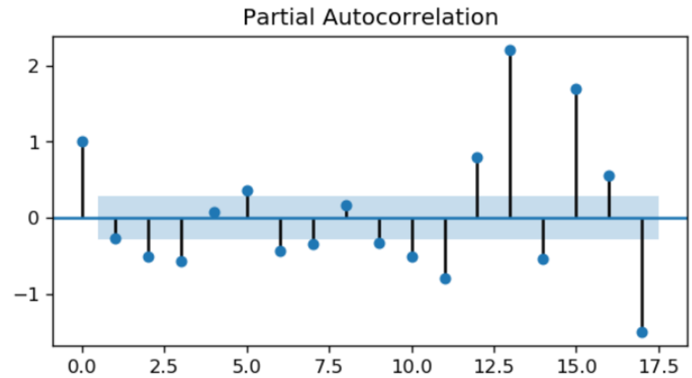


Here, 1st order Differencing gives us stationary data so the d-value is set to 1.

q - VALUE (AUTOCORRELATION)



p - VALUE (PARTIAL AUTOCORRELATION)



ARIMA

From ACF, PACF plots and differencing we get the order as (4,1,0) for the ARIMA model.

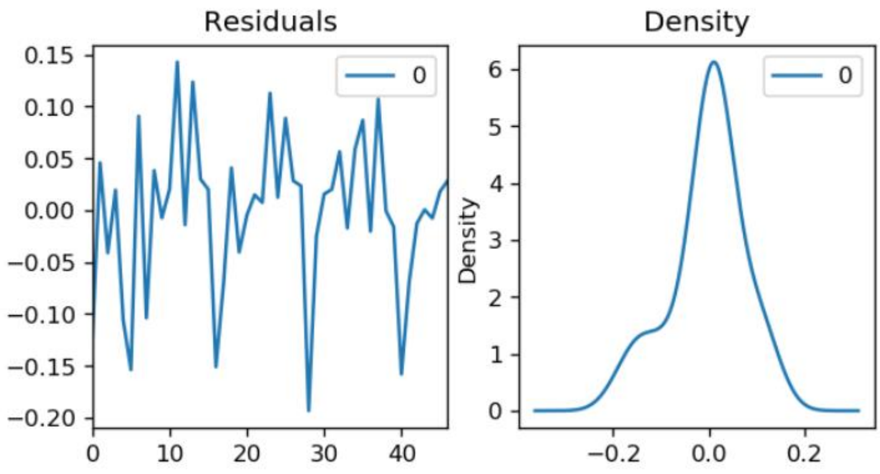
```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                   47
Model:                 ARIMA(3, 1, 0)   Log Likelihood                  55.281
Method:                       css-mle   S.D. of innovations              0.073
Date:                Wed, 15 Jan 2020   AIC                           -100.561
Time:                        12:49:46   BIC                            -91.310
Sample:                             1   HQIC                           -97.080

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0048      0.004      1.256      0.216      -0.003       0.012
ar.L1.D.y     -0.6608      0.124     -5.348      0.000      -0.903      -0.419
ar.L2.D.y     -0.6935      0.115     -6.049      0.000      -0.918      -0.469
ar.L3.D.y     -0.5266      0.120     -4.371      0.000      -0.763      -0.290
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            0.0385            -1.1665j            1.1672           -0.2447
AR.2            0.0385            +1.1665j            1.1672            0.2447
AR.3           -1.3939            -0.0000j            1.3939           -0.5000
-----------------------------------------------------------------------------
```

RMSE: 188599

Residual Diagnosis

Let’s check how the Residuals are distributed.



There is a pattern in residuals and variance is not constant. The density plot is slightly left-skewed.

AUTO ARIMA

AUTOARIMA returns the best ARIMA model after trying all the possible parameters (within the constraints provided) and returns the model with the lowest AIC or BIC.

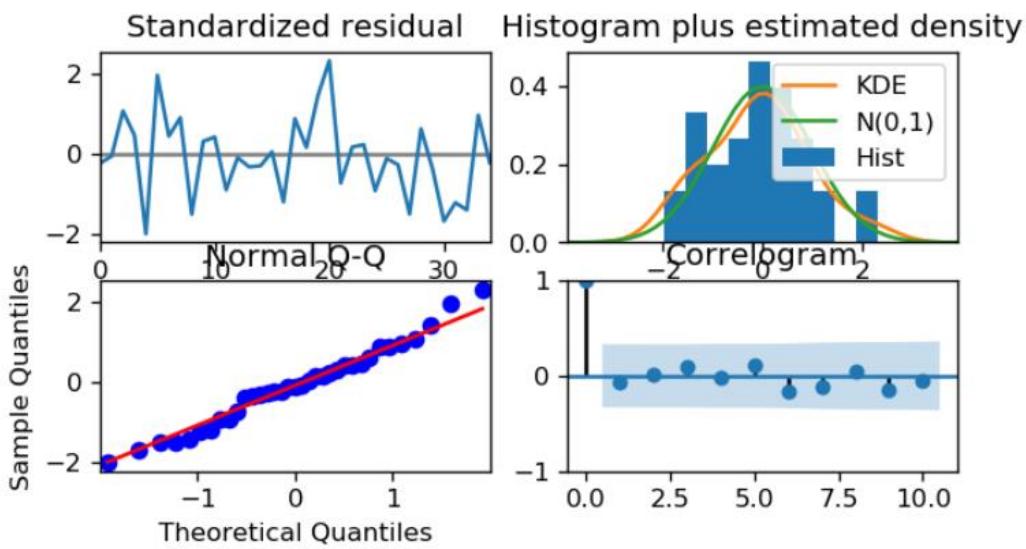
Statespace Model Results

Dep. Variable:	y	No. Observations:	48
Model:	SARIMAX(1, 1, 1)x(0, 1, 0, 12)	Log Likelihood	73.493
Date:	Wed, 15 Jan 2020	AIC	-136.987
Time:	12:50:14	BIC	-129.210
Sample:	0	HQIC	-134.302
	- 48		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0107	0.007	1.517	0.129	-0.003	0.024
drift	-0.0004	0.000	-1.682	0.093	-0.001	6.42e-05
ar.L1	-0.4953	0.252	-1.962	0.050	-0.990	-0.001
ma.L1	-0.6210	0.246	-2.528	0.011	-1.102	-0.139
sigma2	0.0008	0.000	4.028	0.000	0.000	0.001

Ljung-Box (Q):	nan	Jarque-Bera (JB):	0.49
Prob(Q):	nan	Prob(JB):	0.78
Heteroskedasticity (H):	0.65	Skew:	0.23
Prob(H) (two-sided):	0.47	Kurtosis:	3.35

Residual Diagnosis



- The mean of residuals is zero except for a pattern that cannot be captured.
- The residuals are normally distributed.
- The autocorrelation in the correlogram is within the significance level.
- The residuals in QQ plot is roughly a straight line which means that it follows a normal distribution.

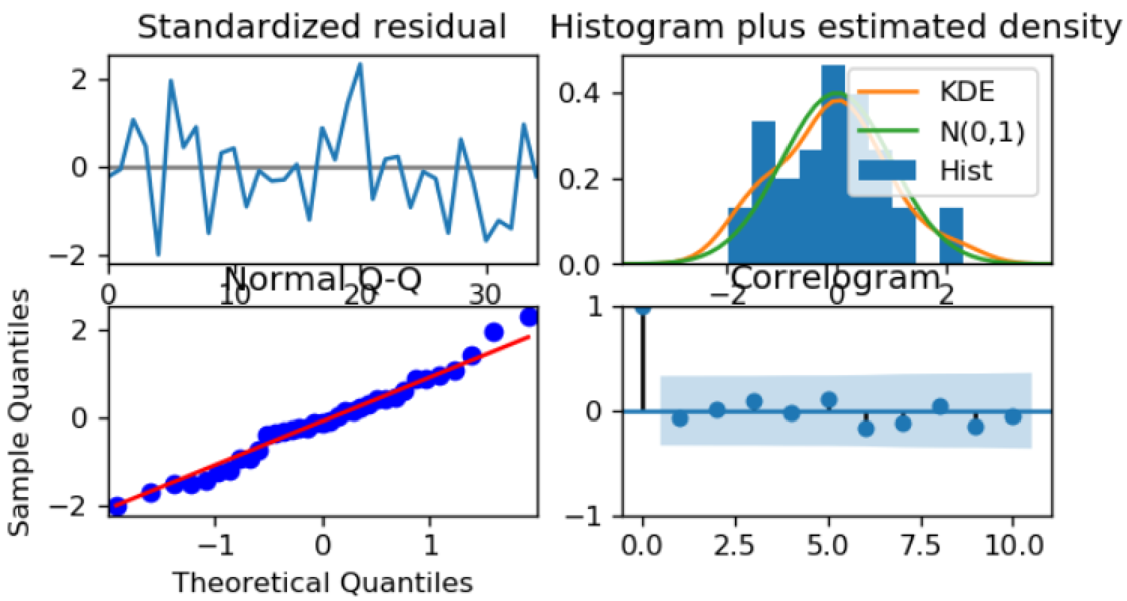
SARIMA

From AUTOARIMA, the order for SARIMA is (1,1,1)x(0,1,0,12)

```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                   48
Model:             SARIMAX(1, 1, 1)x(0, 1, 0, 12)   Log Likelihood                  72.316
Date:                            Wed, 15 Jan 2020   AIC                           -138.631
Time:                                    12:49:57   BIC                           -133.965
Sample:                                         0   HQIC                          -137.026
                                             - 48
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5383      0.226     -2.384      0.017      -0.981      -0.096
ma.L1         -0.4827      0.211     -2.288      0.022      -0.896      -0.069
sigma2         0.0009      0.000      3.830      0.000       0.000       0.001
===================================================================================
Ljung-Box (Q):                         nan   Jarque-Bera (JB):                  0.32
Prob(Q):                               nan   Prob(JB):                          0.85
Heteroskedasticity (H):               0.81   Skew:                              0.21
Prob(H) (two-sided):                  0.72   Kurtosis:                          2.78
===================================================================================
```

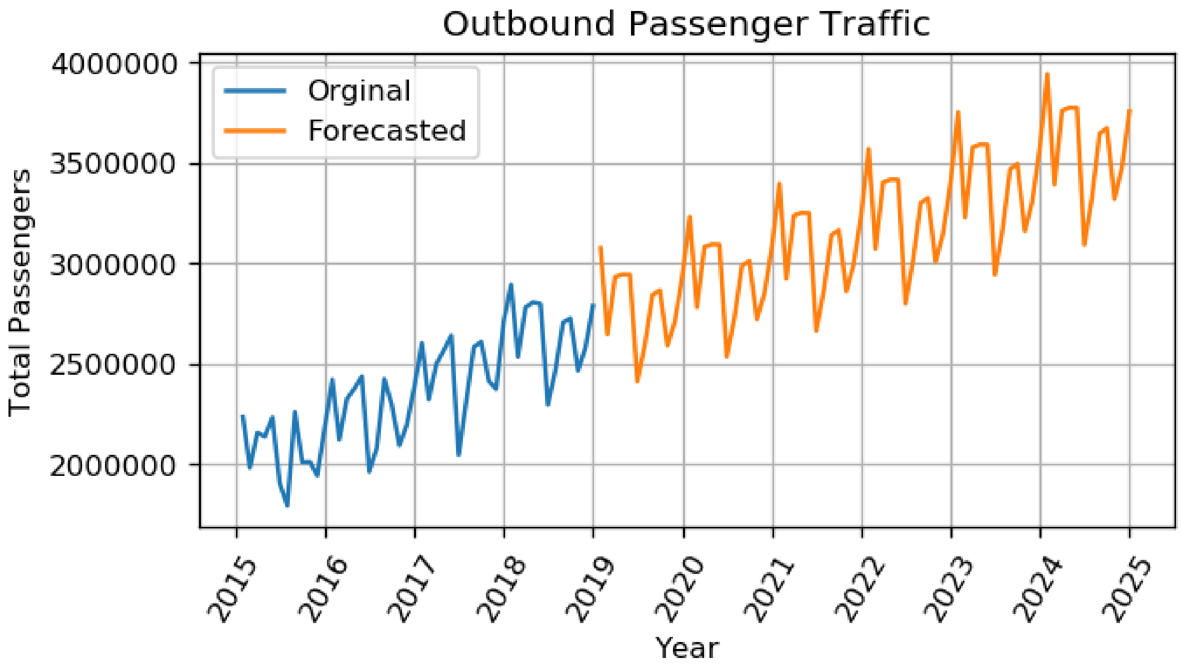
RMSE: 28758

Residual Diagnosis



- The mean of residuals is zero except for a pattern that cannot be captured.
- The residuals are normally distributed.
- The autocorrelation in the correlogram is within the significance level.
- The residuals in QQ plot is roughly a straight line which means that it follows a normal distribution.

Forecast



From the above graph, it can be observed that by 2025 the outbound passenger freight will reach up to 3.8 Million.

10.3 Forecast Model for Inbound Freight Traffic

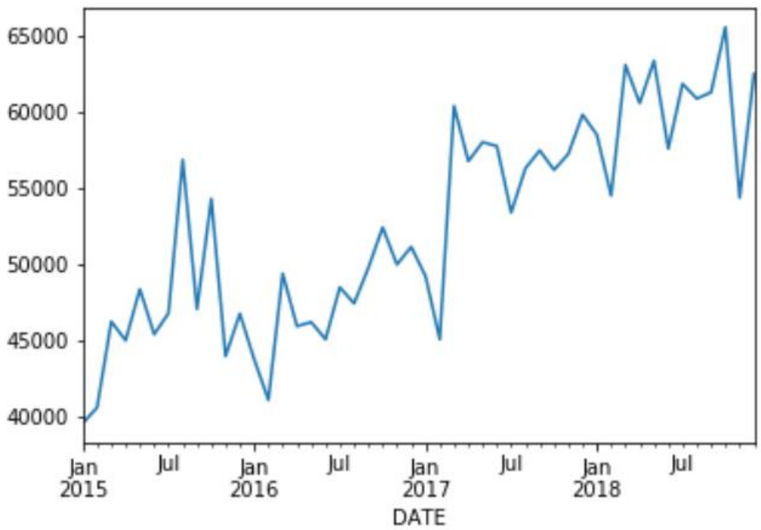
Importing Data

Here’s a look of our sample data.

YEAR	MONTH	QUARTER	AIRLINE_NAME	CARRIER_TYPE	PASSENGERS_TO_INDIA	PASSENGERS_FROM_INDI	FREIGHT_TO_INDIA	FREIGHT_FROM_INDIA	DATE	TOTAL_PASSENGERS	TOTAL_FREIGHT
2015	1	Q1	AIR INDIA	DOMESTIC	258876	274220	3320.626	4186.302	2015-01-31 00:00:00	533096	7506.928
2015	2	Q1	AIR INDIA EXPRI	DOMESTIC	95581	116600	0	0	2015-01-31 00:00:00	212181	0
2015	3	Q1	INDIGO	DOMESTIC	68112	74212	320	1812	2015-01-31 00:00:00	142324	2132
2015	4	Q2	JET AIRWAYS	DOMESTIC	320853	332116	4173.874	5383.515	2015-01-31 00:00:00	652969	9557.389
2015	5	Q2	SPICEJET	DOMESTIC	37882	42468	0	115.68	2015-01-31 00:00:00	80350	115.68
2015	6	Q2	AEROFLOT	FOREIGN	5088	6901	7.581	199.696	2015-01-31 00:00:00	11989	207.277
2015	7	Q3	AEROLOGIC	FOREIGN	0	0	1.409	1.235	2015-01-31 00:00:00	0	2.644
2015	8	Q3	AIR ARABIA	FOREIGN	77405	79096	191.549	1096.201	2015-01-31 00:00:00	156501	1287.75
2015	9	Q3	AIR ASIA BERHA	FOREIGN	25906	27950	33.524	307.431	2015-01-31 00:00:00	53856	340.955
2015	10	Q4	AIR AUSTRAL	FOREIGN	0	0	0	0	2015-01-31 00:00:00	0	0
2015	11	Q4	AIR CHINA	FOREIGN	4004	4047	159.159	191.009	2015-01-31 00:00:00	8051	350.168
2015	12	Q4	AIR FRANCE	FOREIGN	20204	22839	726.368	1365.611	2015-01-31 00:00:00	43043	2091.979
2016	1	Q1	AIR MANAS	FOREIGN	0	0	0	0	2015-01-31 00:00:00	0	0
2016	2	Q1	AIR MAURITIUS	FOREIGN	5987	7219	27.508	263.285	2015-01-31 00:00:00	13206	290.793
2016	3	Q1	AIR ASTANA	FOREIGN	2608	3080	0	25.509	2015-01-31 00:00:00	5688	25.509
2016	4	Q2	AIR SEYCHELLES	FOREIGN	1253	1300	0.514	1.635	2015-01-31 00:00:00	2553	2.149

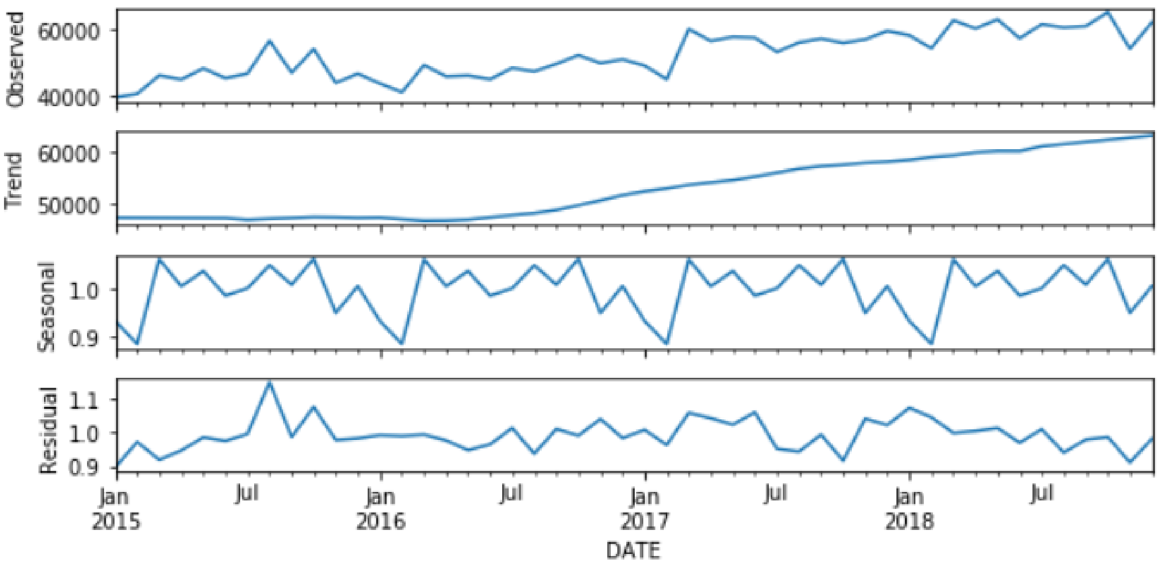
Original Series

At the beginning it is worth looking at the general cargo services operations. Below graph shows freight charges to India over time.

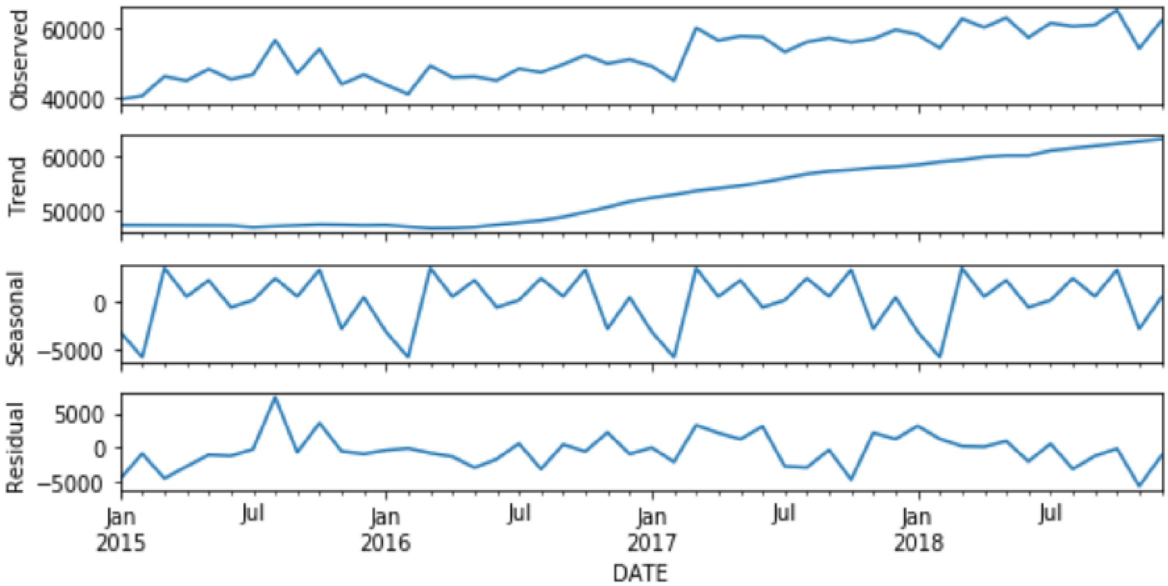


Decomposition

Multiplicative Decompose



Additive Decompose

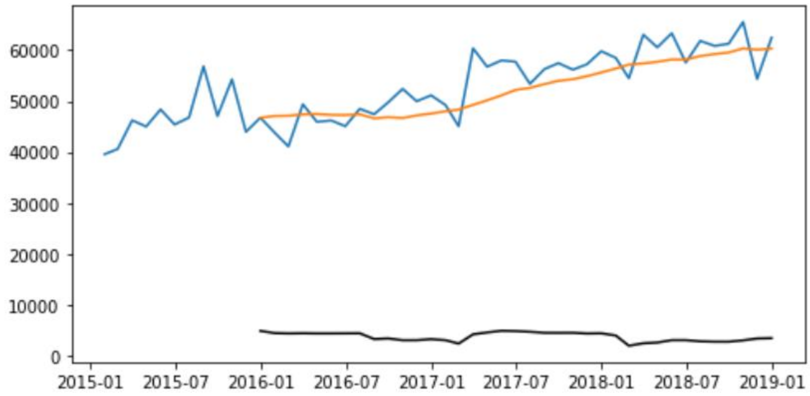


The model seems to have a 12-month seasonal trend which makes sense as usually every year a holiday period is more busy at airports.

Test for Stationarity

Let’s plot the rolling mean and rolling standard deviation having window size of 12-months over original data.

```
ADF Statistic: -1.364093
p-value: 0.599364
```

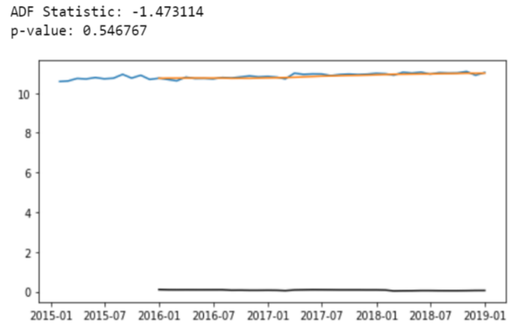


- Blue line represents our original data
- Orange line represents rolling mean and Black line represents rolling standard deviation.
- We can see that the mean and standard deviation are not constant over time. Hence the series is not stationary.
- Here p-value>0.05, so we reject the Null Hypothesis and conclude that the series is non-stationary.

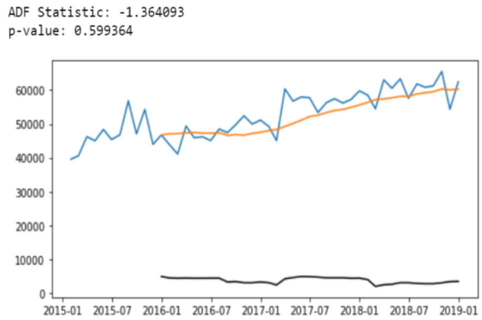
Making Time Series Stationary

In order to stationarize time series and to ARMA model, I will transform this time series by taking the 1-month difference.

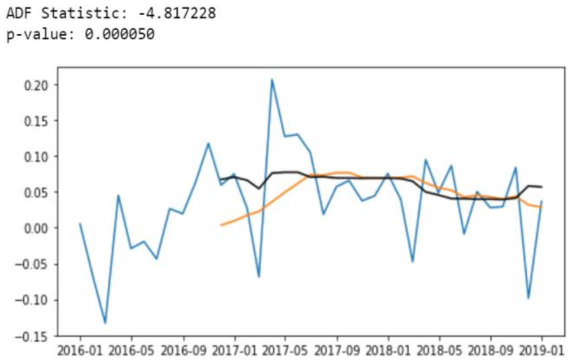
Original Series



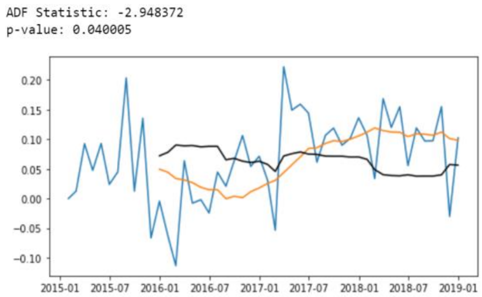
Log Transformation



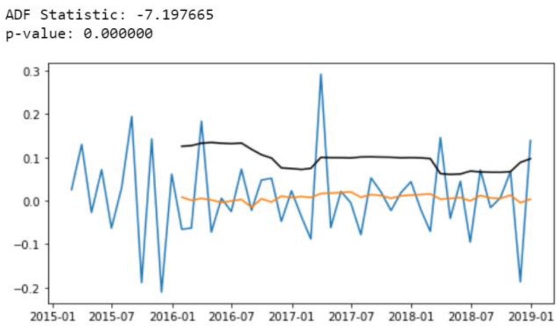
Log Minus Weighted Average



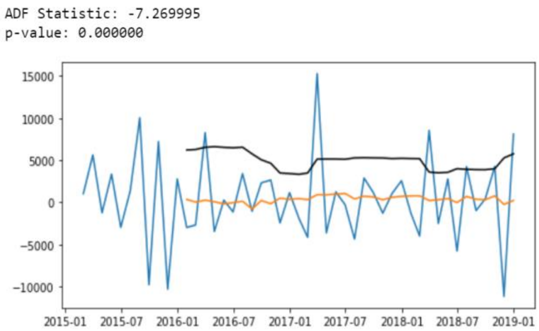
Log Minus Moving Average



Log Differencing



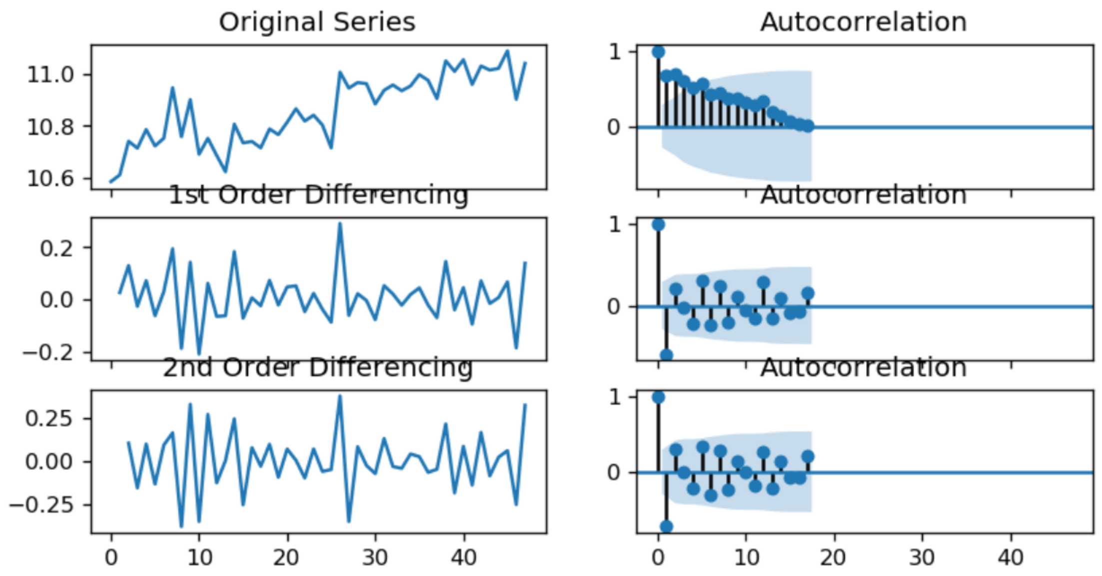
Differencing



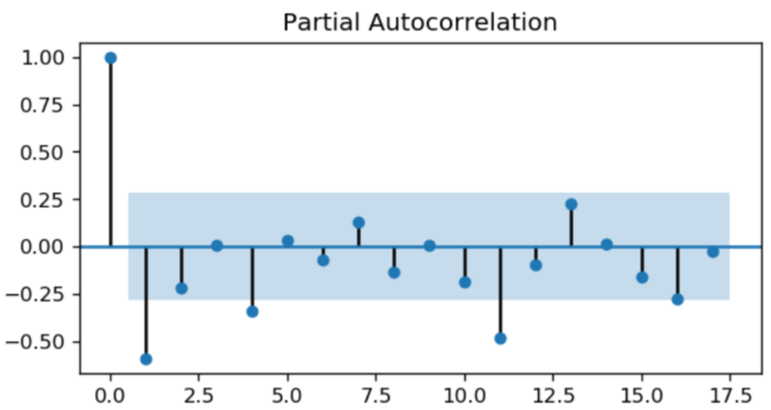
- From the above graphs, we see that Log Differencing and Differencing methods have relatively stationary data having p-value < 0.05.
- Hence we have transformed the time series data and is now stationary.

Finding Model Parameters (p,d,q)

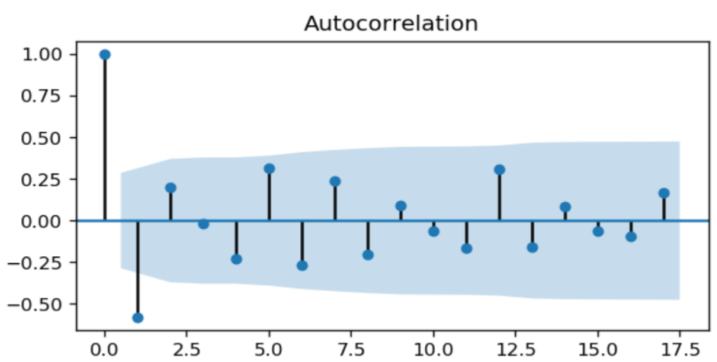
d- VALUE(N-order Differencing)



p - VALUE (PARTIAL AUTOCORRELATION)



q - VALUE (AUTOCORRELATION)



ARIMA Model

From the PACF AND ACF plots we get ARIMA order of (1,1,0) by differencing for 1-month.

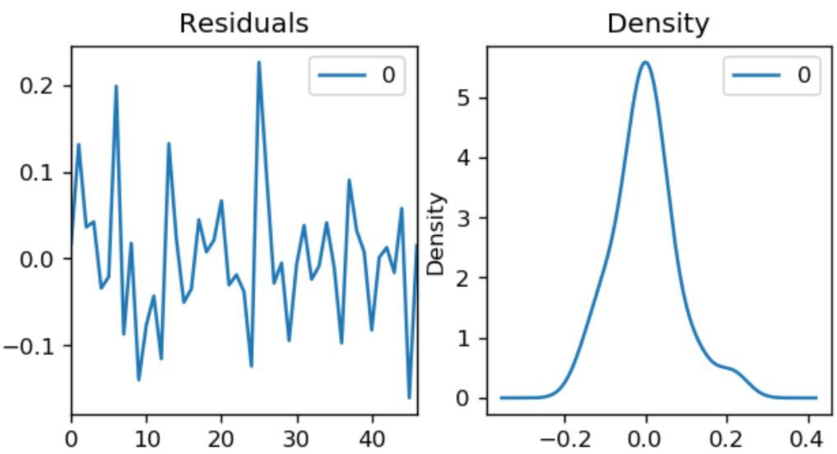
```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                   47
Model:                 ARIMA(1, 1, 0)   Log Likelihood                  52.877
Method:                       css-mle   S.D. of innovations              0.078
Date:                Mon, 13 Jan 2020   AIC                            -99.754
Time:                        15:42:28   BIC                            -94.203
Sample:                             1   HQIC                           -97.665

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0085      0.007      1.182      0.243      -0.006       0.023
ar.L1.D.y     -0.5911      0.117     -5.039      0.000      -0.821      -0.361
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -1.6916           +0.0000j            1.6916            0.5000
-----------------------------------------------------------------------------
```

RMSE = 7788.829529626096

Residual Diagnosis

Let’s check how the Residuals are distributed.



AUTO ARIMA

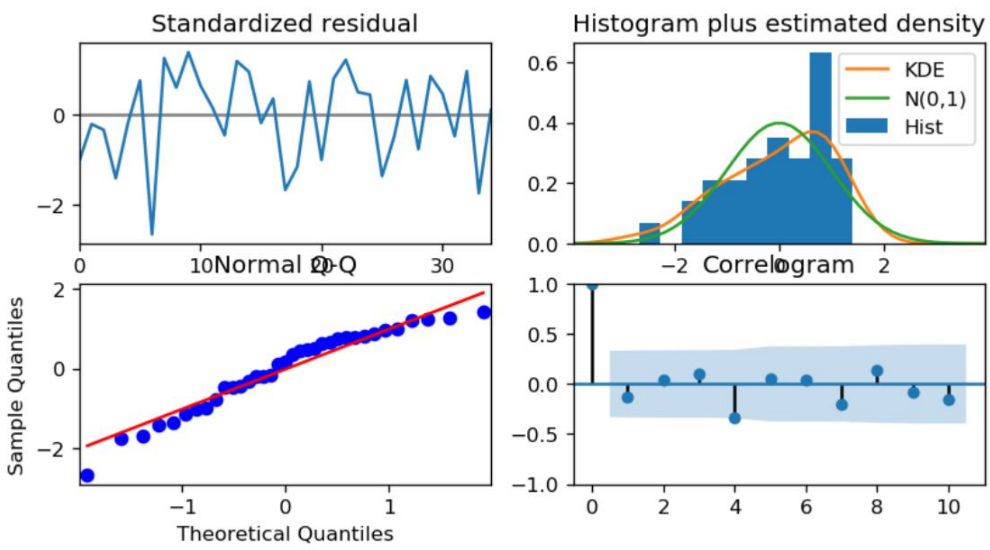
AUTOARIMA returns the best ARIMA model after trying all the possible parameters (within the constraints provided) and returns the model with the lowest AIC or BIC.

Dep. Variable:	y	No. Observations:	48
Model:	SARIMAX(1, 1, 0)x(1, 1, 0, 12)	Log Likelihood	42.062
Date:	Thu, 26 Dec 2019	AIC	-74.123
Time:	16:37:10	BIC	-66.347
Sample:	0	HQIC	-71.439
	- 48		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0416	0.063	0.666	0.506	-0.081	0.164
drift	-0.0013	0.002	-0.741	0.458	-0.005	0.002
ar.L1	-0.6228	0.180	-3.467	0.001	-0.975	-0.271
ar.S.L12	-0.3255	0.218	-1.493	0.136	-0.753	0.102
sigma2	0.0050	0.002	2.956	0.003	0.002	0.008

Ljung-Box (Q):	nan	Jarque-Bera (JB):	2.67
Prob(Q):	nan	Prob(JB):	0.26
Heteroskedasticity (H):	0.58	Skew:	-0.66
Prob(H) (two-sided):	0.37	Kurtosis:	2.66

Residual Diagnosis



- The mean of residuals is zero except for a pattern that cannot be captured.
- The residuals are normally distributed.
- The autocorrelation in the correlogram is within the significance level.
- The residuals in QQ plot is roughly a straight line which means that it follows a normal distribution.

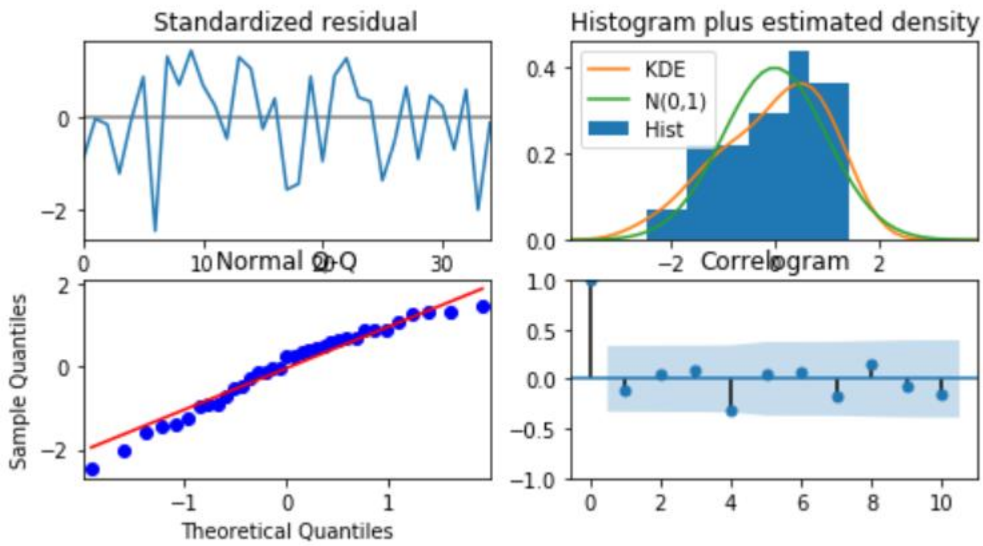
SARIMA

From AUTOARIMA, the order for SARIMA is (1,1,0)x(1,1,0,12)

```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                   48
Model:             SARIMAX(1, 1, 0)x(1, 1, 0, 12)   Log Likelihood                  41.926
Date:                            Wed, 15 Jan 2020   AIC                            -77.852
Time:                                    15:52:32   BIC                            -73.186
Sample:                                         0   HQIC                           -76.241
                                             - 48
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6318      0.161     -3.919      0.000      -0.948      -0.316
ar.S.L12      -0.4239      0.189     -2.239      0.025      -0.795      -0.053
sigma2         0.0049      0.001      3.430      0.001       0.002       0.008
===================================================================================
Ljung-Box (Q):                         nan   Jarque-Bera (JB):                 2.16
Prob(Q):                               nan   Prob(JB):                         0.34
Heteroskedasticity (H):               0.63   Skew:                            -0.55
Prob(H) (two-sided):                  0.44   Kurtosis:                         2.47
===================================================================================
```

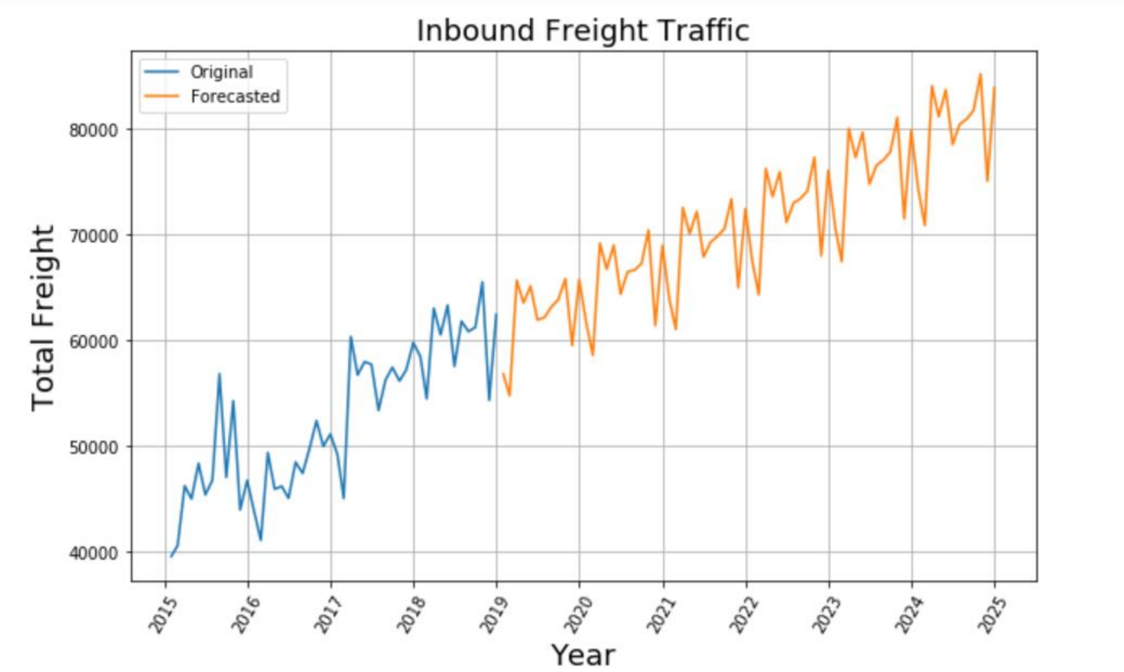
RMSE= 1926.51

Residual Diagnosis



- The mean of residuals is zero except for a pattern that cannot be captured.
- The residuals are normally distributed.
- The autocorrelation in the correlogram is within the significance level.
- The residuals in QQ plot is roughly a straight line which means that it follows a normal distribution.

Forecast

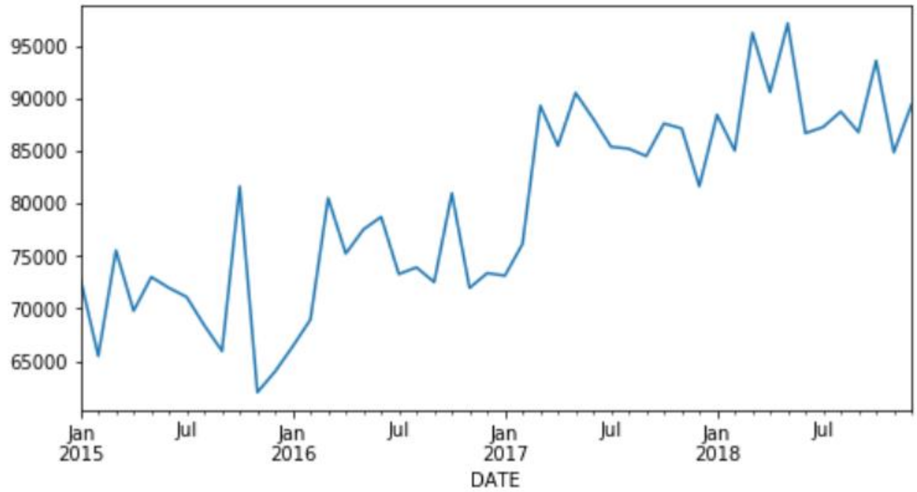


The total Freight traffic to India from other countries is expected to reach up to 10 Million by 2025**.**

10.3 Forecast Model for Outbound Freight Traffic

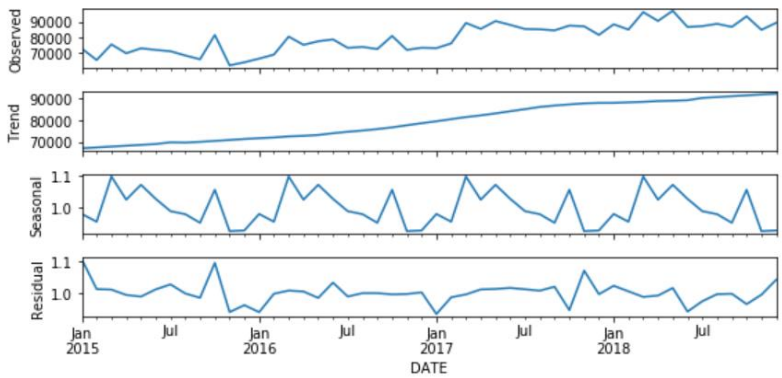
Original Series

At the beginning it is worth looking at the general cargo services operations. Below graph shows freight charges from India over time.

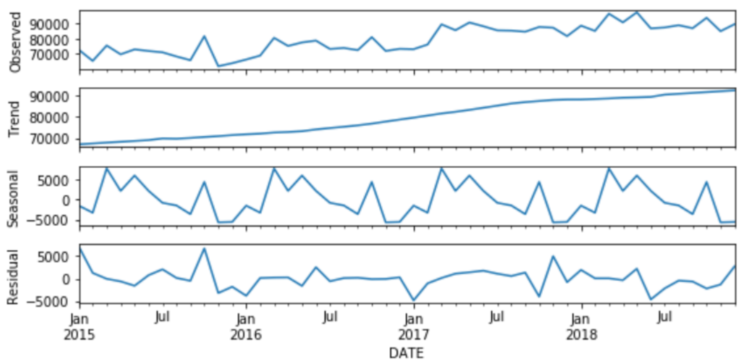


Decomposition

Multiplicative Decompose



Additive Decompose

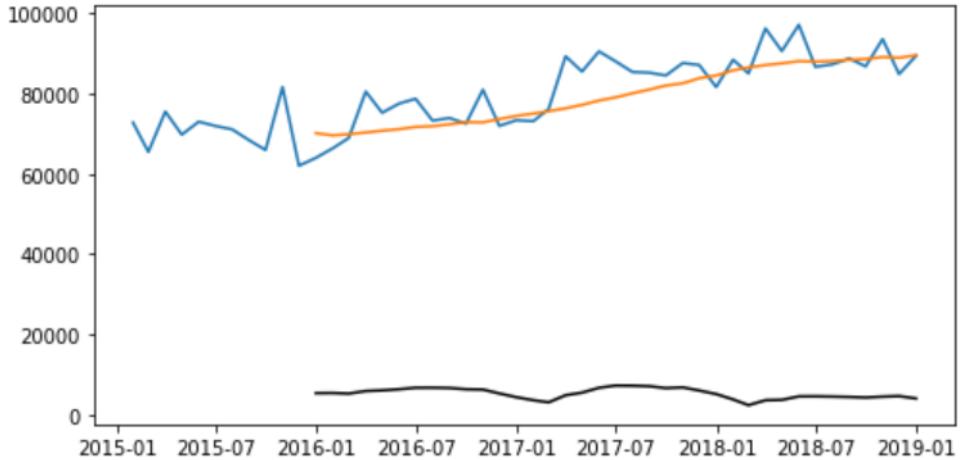


The model seems to have a 12-month seasonal trend which makes sense as usually every year a holiday period is more busy at airports.

Test for Stationarity

Let’s plot the rolling mean and rolling standard deviation having window size of 12-months over original data.

```
ADF Statistic: -0.826387
p-value: 0.811248
```

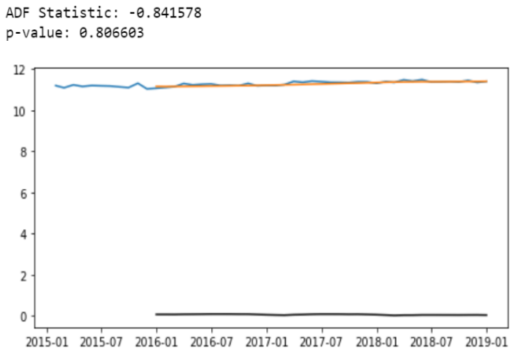


- Blue line represents our original data
- Orange line represents rolling mean and Black line represents rolling standard deviation.
- We can see that the mean and standard deviation are not constant over time. Hence the series is not stationary.
- Here p-value>0.05, so we reject the Null Hypothesis and conclude that the series is non-stationary.

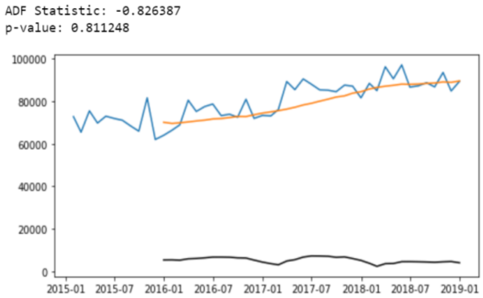
Making Time Series Stationary

In order to stationarize time series and to ARMA model, I will transform this time series by taking the 1-month difference.

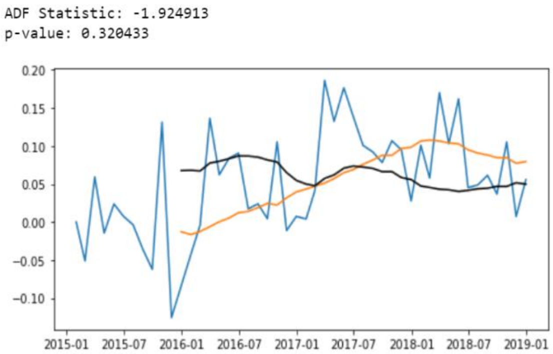
Original Series



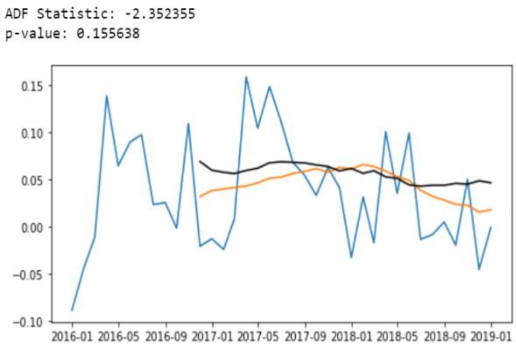
Log Transformation



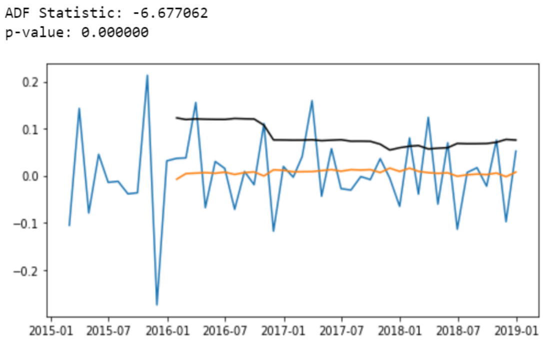
Log Minus Weighted Average



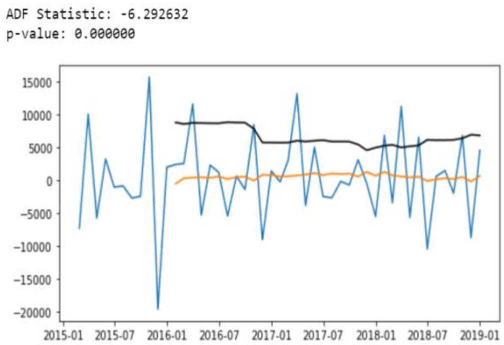
Log Minus Moving Average



Log Differencing



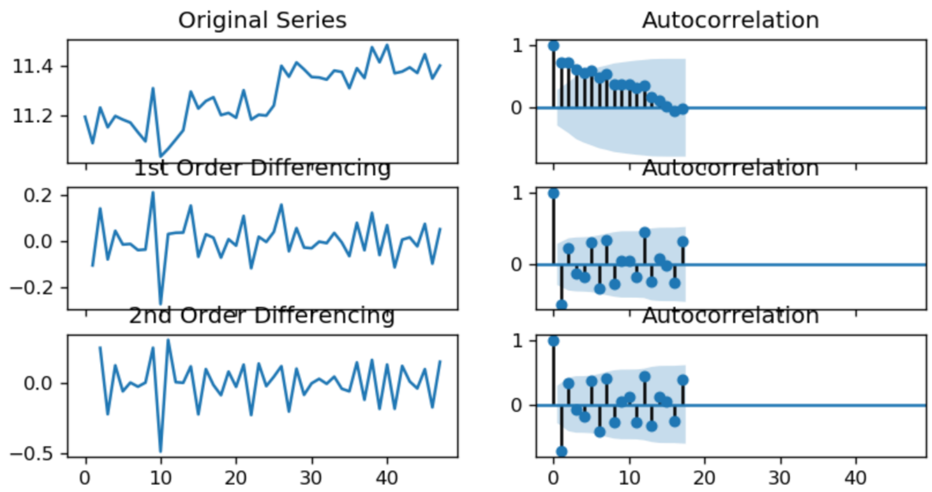
Differencing



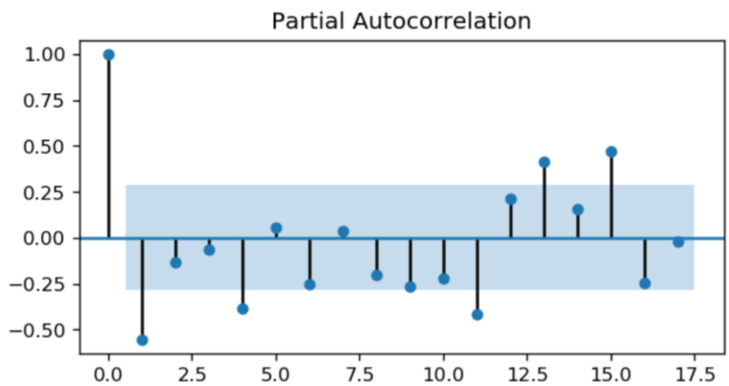
From the above graphs, we see that Log Differencing and Differencing methods have relatively stationary data having $p\text{-value} < 0.05$.
Hence we have transformed the time series data and is now stationary.

Finding Model Parameters(p,d,q)

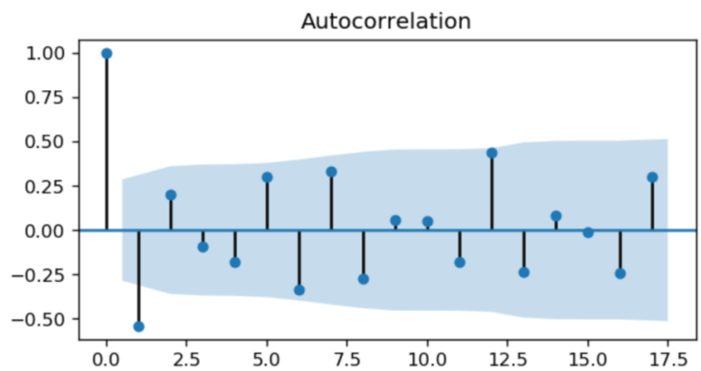
d- VALUE(N-order Differencing)



p - VALUE (PARTIAL AUTOCORRELATION)



q - VALUE (AUTOCORRELATION)



ARIMA

From the PACF AND ACF plots we get ARIMA order of (1,1,0) by differencing for 1-month.

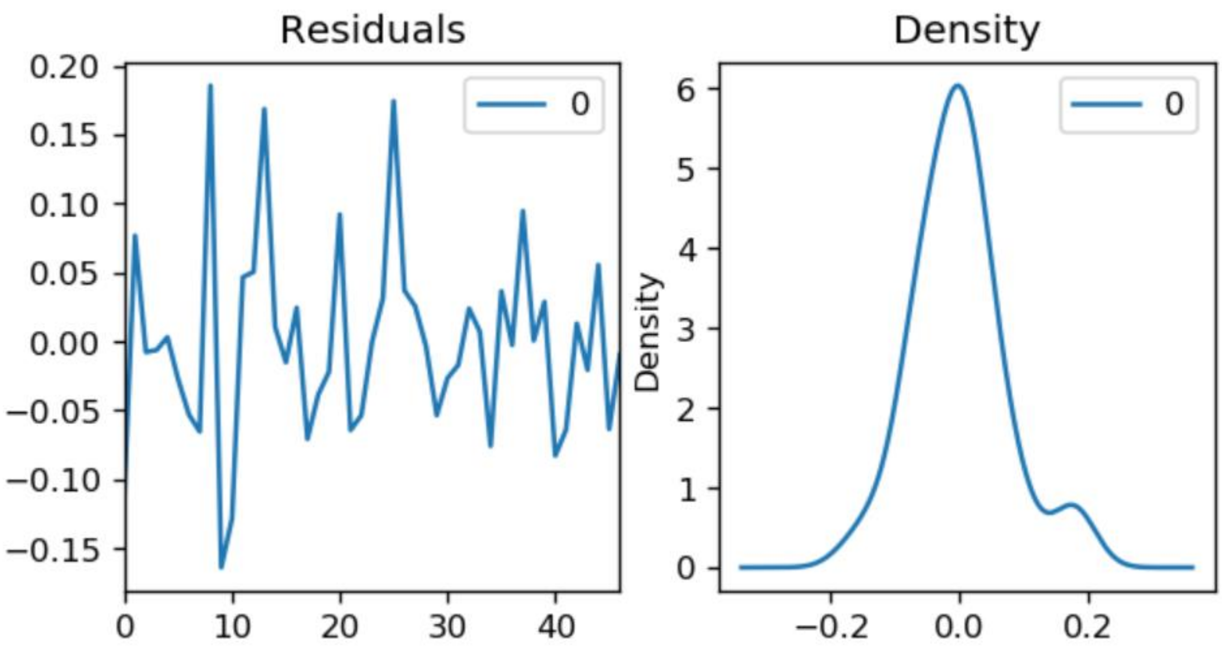
```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                   47
Model:                 ARIMA(1, 1, 0)   Log Likelihood                  58.177
Method:                       css-mle   S.D. of innovations              0.070
Date:                Mon, 13 Jan 2020   AIC                           -110.353
Time:                        12:16:47   BIC                           -104.803
Sample:                             1   HQIC                          -108.265

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0049      0.007      0.734      0.467      -0.008       0.018
ar.L1.D.y     -0.5542      0.122     -4.555      0.000      -0.793      -0.316
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -1.8045           +0.0000j            1.8045            0.5000
-----------------------------------------------------------------------------
```

RMSE = 6347.60

Residual Diagnosis

Let’s check how the Residuals are distributed.



AUTO ARIMA

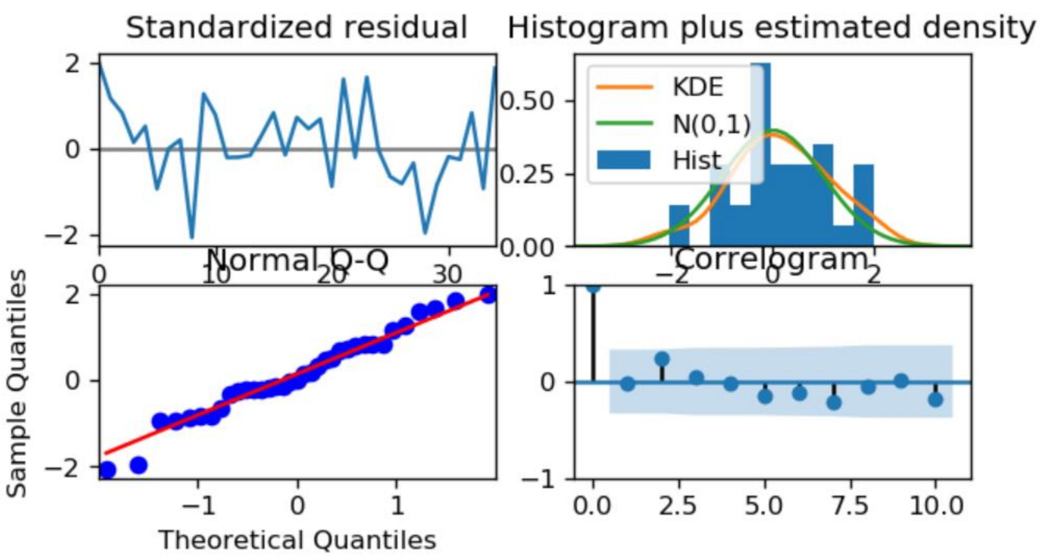
AUTOARIMA returns the best ARIMA model after trying all the possible parameters (within the constraints provided) and returns the model with the lowest AIC or BIC.

Dep. Variable:	y	No. Observations:	48
Model:	SARIMAX(0, 1, 1)x(0, 1, 0, 12)	Log Likelihood	56.961
Date:	Mon, 13 Jan 2020	AIC	-105.923
Time:	12:21:40	BIC	-99.701
Sample:	0	HQIC	-103.775
	- 48		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0217	0.007	3.225	0.001	0.009	0.035
drift	-0.0007	0.000	-3.497	0.000	-0.001	-0.000
ma.L1	-0.9308	0.190	-4.889	0.000	-1.304	-0.558
sigma2	0.0023	0.001	3.427	0.001	0.001	0.004

Ljung-Box (Q):	nan	Jarque-Bera (JB):	0.15
Prob(Q):	nan	Prob(JB):	0.93
Heteroskedasticity (H):	0.98	Skew:	-0.14
Prob(H) (two-sided):	0.97	Kurtosis:	2.84

Residual Diagnosis



- The mean of residuals is zero except for a pattern that cannot be captured.
- The residuals are normally distributed.
- The autocorrelation in the correlogram is within the significance level. The residuals in QQ plot is roughly a straight line which means that it follows a normal distribution.

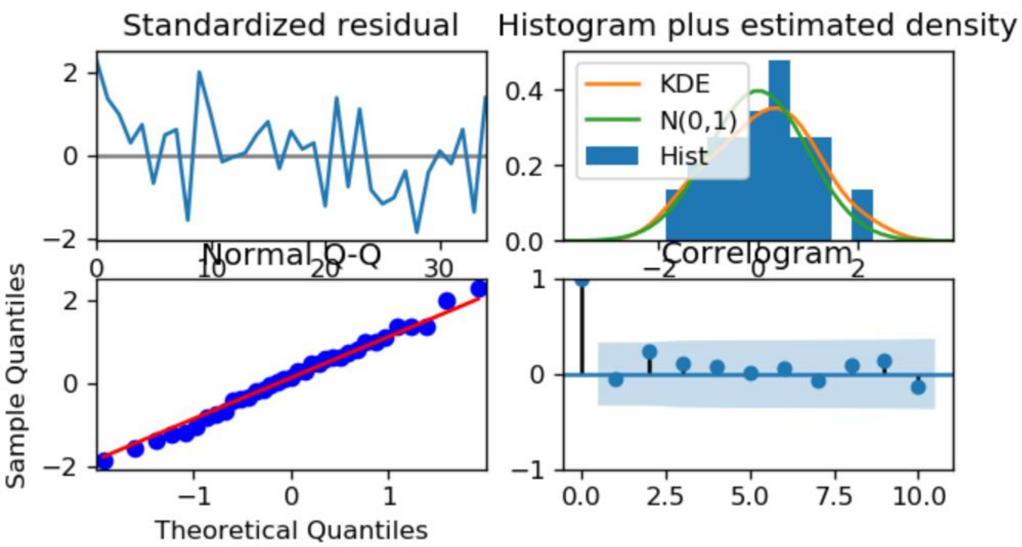
SARIMA

From AUTOARIMA, the order for SARIMA is (0,1,1)x(0,1,0,12)

```
==========================================================================================
Dep. Variable:                                  y   No. Observations:                   48
Model:             SARIMAX(0, 1, 1)x(0, 1, 0, 12)   Log Likelihood                  52.853
Date:                            Mon, 13 Jan 2020   AIC                           -101.705
Time:                                    12:24:54   BIC                            -98.595
Sample:                                         0   HQIC                          -100.631
                                             - 48
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6045      0.176     -3.427      0.001      -0.950      -0.259
sigma2         0.0028      0.001      3.434      0.001       0.001       0.004
===================================================================================
Ljung-Box (Q):                       nan   Jarque-Bera (JB):                 0.41
Prob(Q):                             nan   Prob(JB):                         0.81
Heteroskedasticity (H):             0.71   Skew:                             0.01
Prob(H) (two-sided):                0.57   Kurtosis:                         2.47
===================================================================================
```

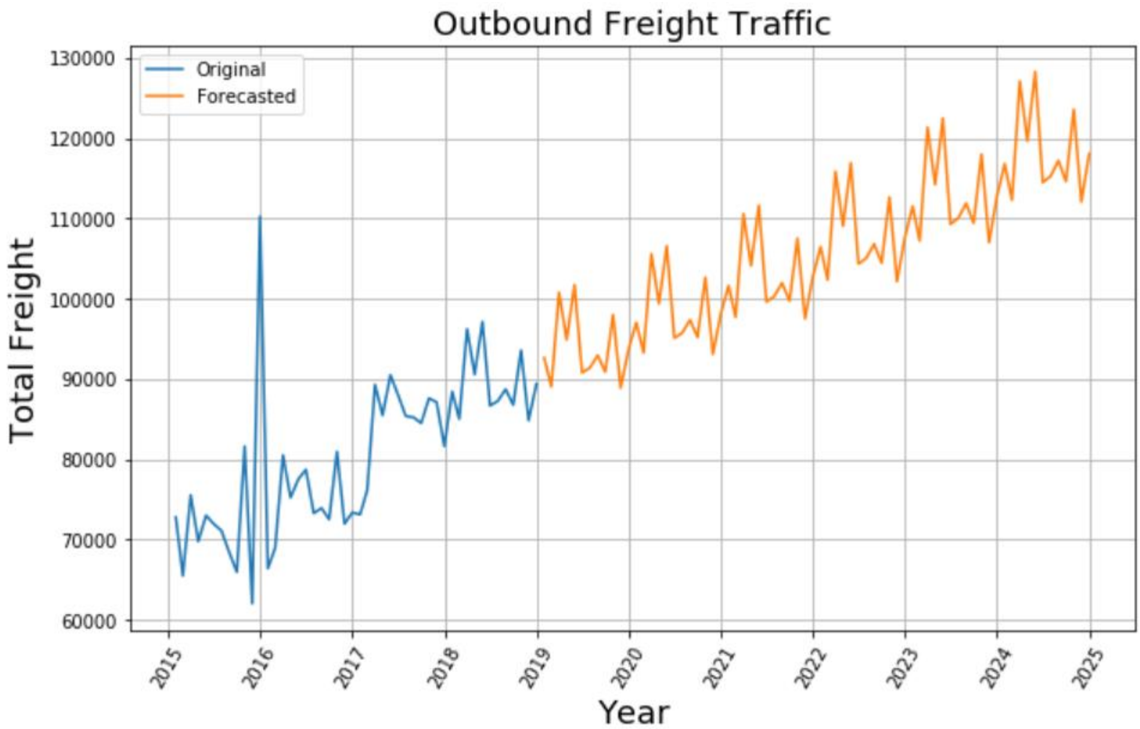
RMSE= 1437.68

Residual Diagnosis



- The mean of residuals is zero except for a pattern that cannot be captured.
- The residuals are normally distributed.
- The autocorrelation in the correlogram is within the significance level. The residuals in QQ plot is roughly a straight line which means that it follows a normal distribution

Forecast



The total Freight traffic from India to other countries is expected to reach up to 13 Million by 2025

Chapter 11: Model Evaluation

Summary of Models:

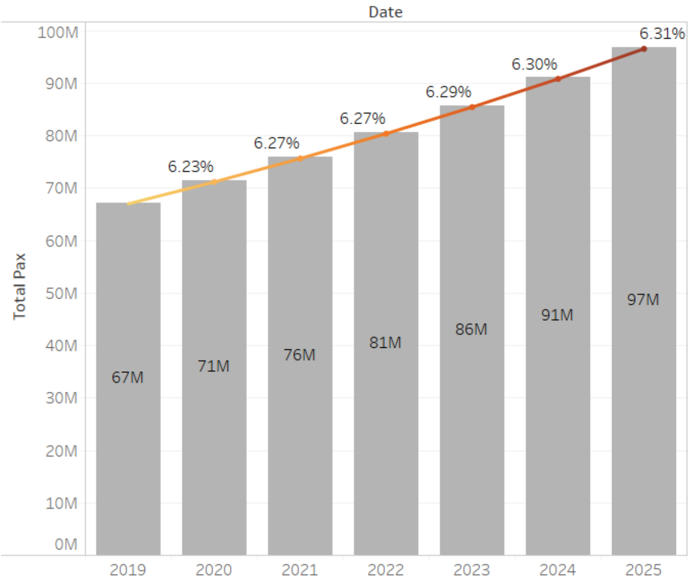
	Model	AIC	RMSE
Forecast Model for Inbound Passenger Traffic	ARIMA(3,1,0)	-110.934	245218
	SARIMA(0,1,1)x(1,1,0,12)	-127.141	17362
Forecast Model for Outbound Passenger Traffic	ARIMA(3,1,0)	-100.56	188599
	SARIMA(1,1,0)x(0,1,0,12)	-138.63	28759
Forecast Model for Inbound Freight Traffic	ARIMA(1,1,0)	-99.75	7788.82
	SARIMA(1,1,0)x(1,1,0)	-77.84	1926.51
Forecast Model for Outbound Freight Traffic	ARIMA(1,1,0)	-110.35	6347.60
	SARIMA(0,1,1)x(0,1,0)	-101.70	1437.68

- Forecast model for Inbound Passenger Traffic, ARIMA model gave an RSME of approx. 2.5 Lakhs (Passengers), when added a seasonal component the SARIMA model gave an RMSE of less than 20K (Passengers), which is acceptable for a forecast range of 2-5 Million.
- Forecast model for Outbound Passenger Traffic, ARIMA model gave an RSME of approx. 2 Lakhs (Passengers), when added a seasonal component the SARIMA model gave an RMSE of less than 30K (Passengers), which is acceptable for a forecast range of 2-5 Million.
- Forecast model for Inbound and Outbound Freight Traffic has an RMSE of less than 2K Tonnes on the monthly forecast, which is acceptable for a forecast range of 50K+ tonnes.

Chapter 12: Future Traffic Projections

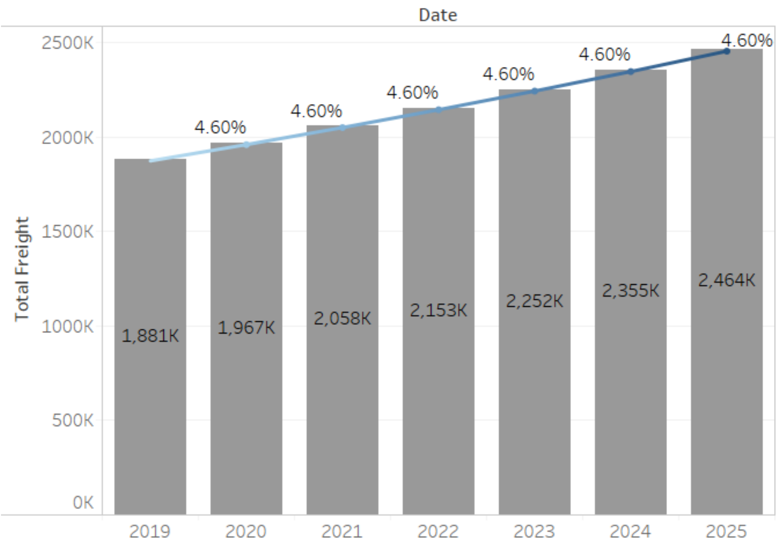
12.1 Total Traffic Projections

Estimated Growth of International Passenger Traffic



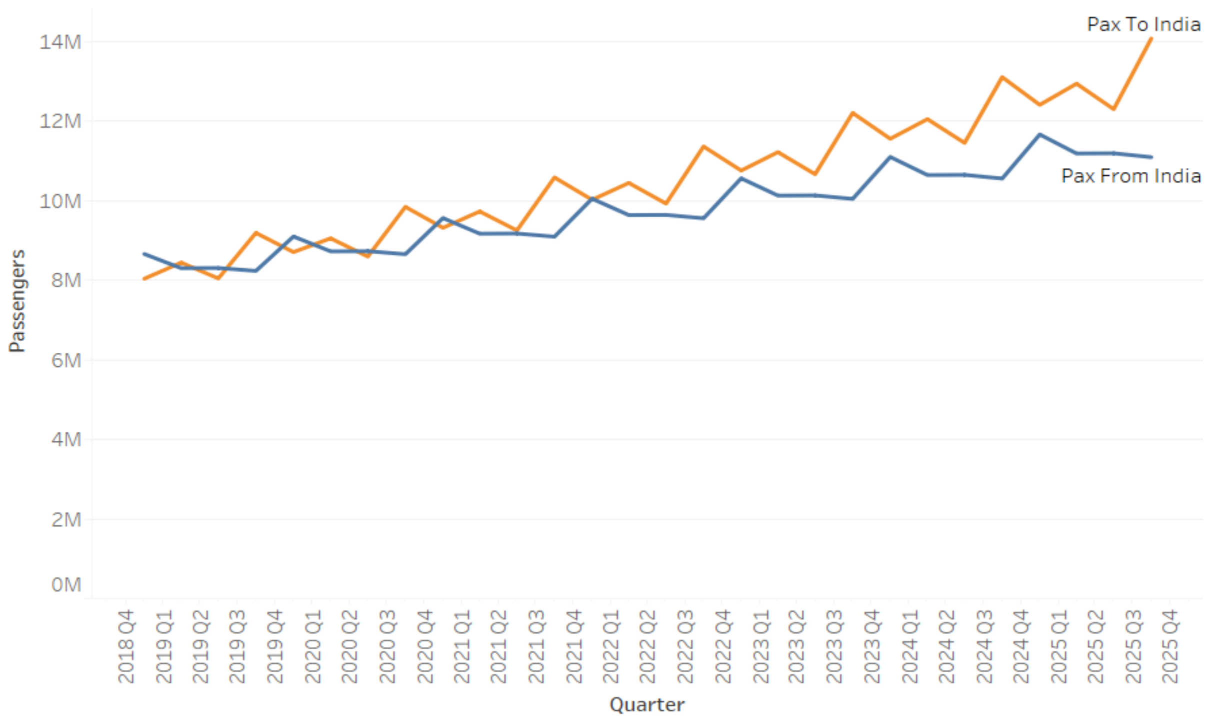
- By 2025 International Passenger Traffic is estimated to grow nearly to 100 Million.
- The International Passenger Traffic is expected to grow an average rate of 6.3%.
- The average revenue per International Passenger is Rs1,500 and the Total Revenue from the International Passenger Traffic is estimated to reach up Rs15,000 Cr by 2025.
- India is expected to have a total of 500 Million International Passengers from 2020 – 2025 which includes an Inbound traffic of 300 Million.

Estimated Growth of International Freight Traffic (in Tonnes)



- By 2025 International Freight Traffic is estimated to grow nearly to 2.5 Million Tonnes.
- International Freight Traffic is expected to grow at a rate of 4.6%.

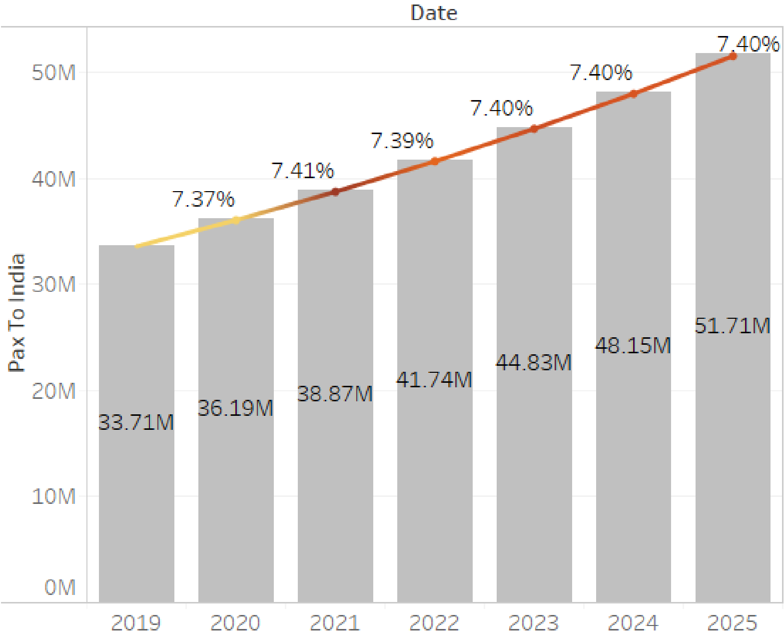
12.2 Projections of Inbound and Outbound Passenger Traffic



The projections show a huge growth of inbound and outbound international passenger traffic till 2025. The pattern observed in the historical data, that the inbound traffic in Q4 of every year is more and outbound traffic in Q1 is more is also observed in the forecast.

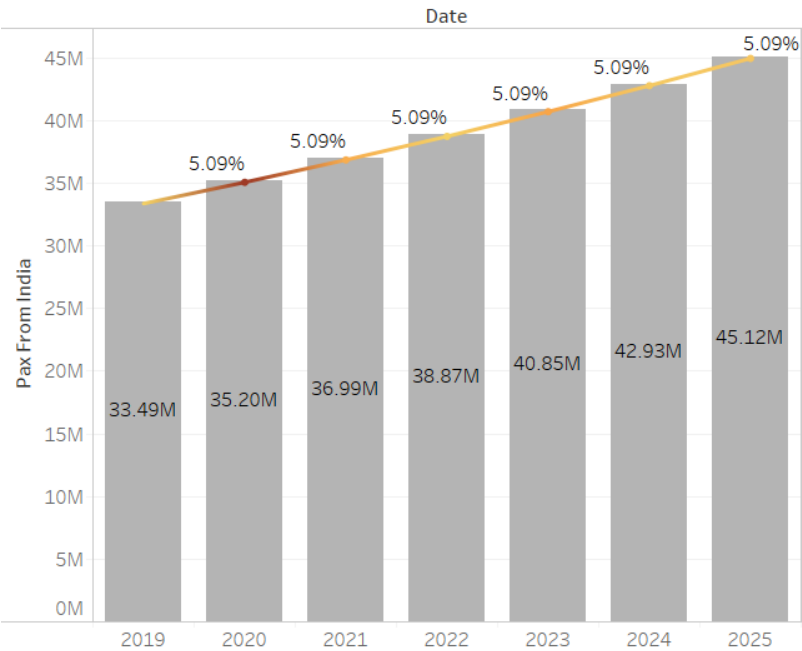
- The inbound international passenger traffic is expected to reach upto 14 Million my Q4 in 2025.
- The outbound traffic will still remain less than 10 Million by Q4 of 2025.
- The growth of inbound traffic in more compared to the outbound traffic.

Growth of Inbound Traffic



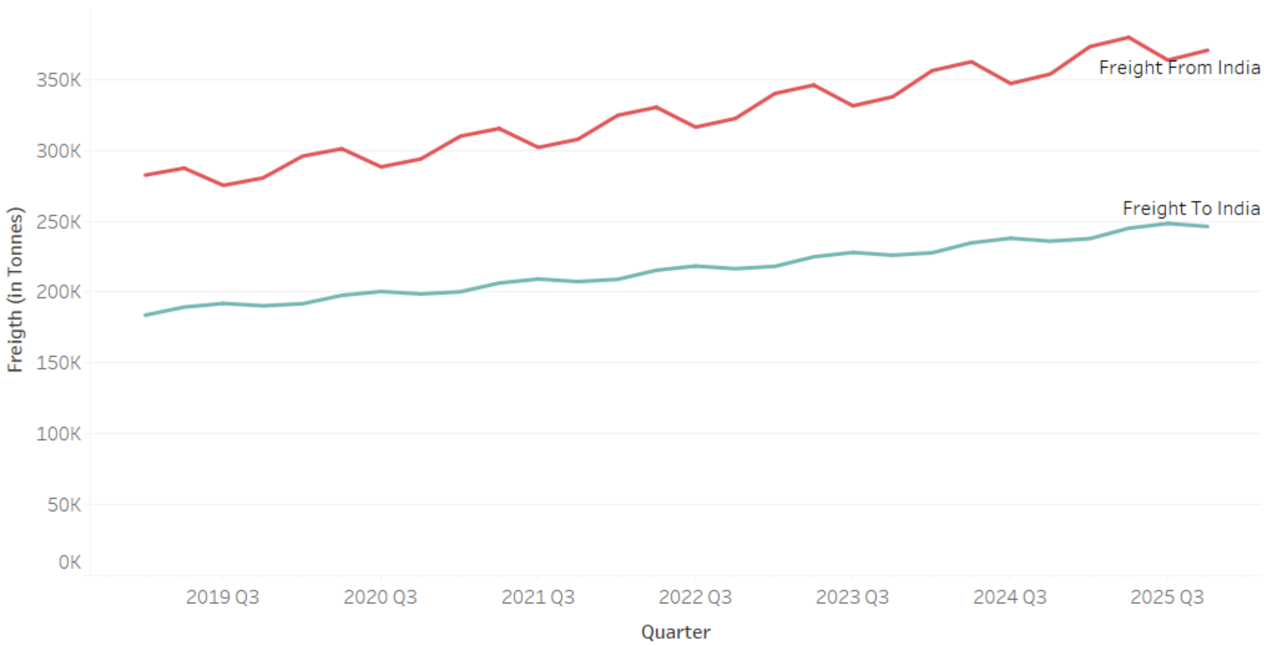
The inbound international passenger traffic in expected to grow at an average of 7.4% every year reaching upto 51 Million by the year 2025.

Growth of Outbound Traffic



The outbound international passenger traffic in expected to grow at an average of 5% every year reaching upto 45 Million by the year 2025.

12.3 Projections of Inbound and Outbound Passenger Traffic



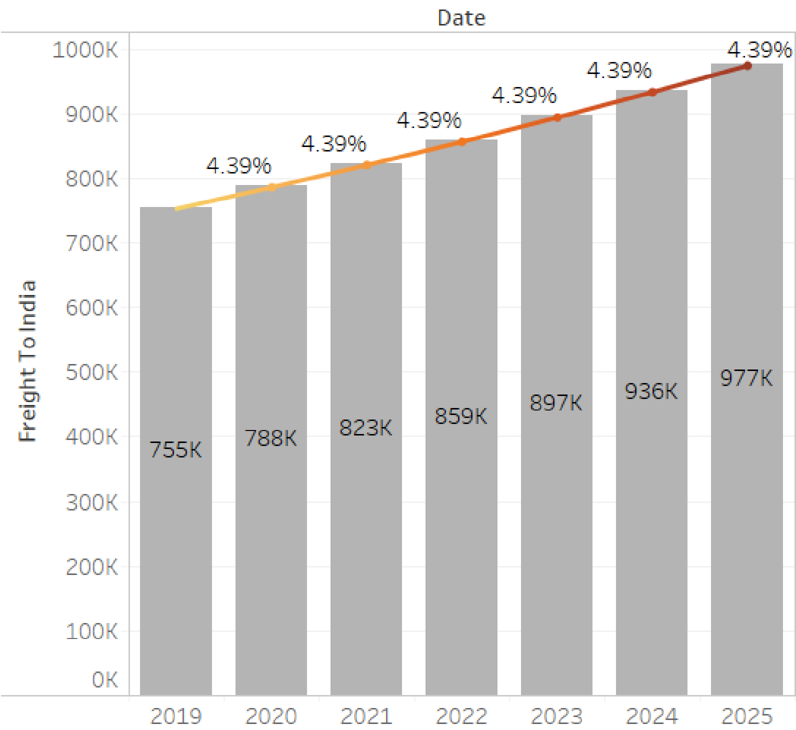
The projections show a growth the inbound and outbound freight traffic, while outbound traffic is more than then inbound traffic. The projected outbound traffic has a pattern that seems to decrease during Q4 and increases during Q1.

- The outbound traffic is expected to reach almost 400K Tonnes in Q4 2015.
- The inbound traffic is projected less than 250K Tonnes in 2025.

The projections for outbound and inbound traffic have a difference of more than 150K Tonnes which suggests than India has more Exports. Increase in Exports suggests a huge growth in the economy of India.

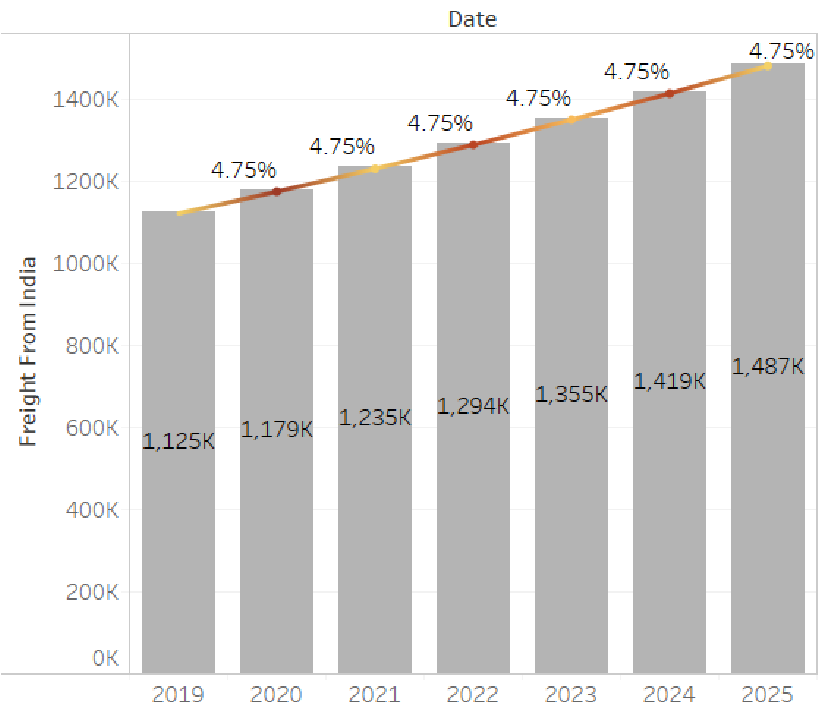
The projections will help us to maintain the Import-Export ratio, while focusing more on the exports. This is useful for the airlines in planning and scheduling outbound cargo services as it is more than the inbound freight traffic.

Growth of Inbound Freight Traffic



The inbound international freight traffic in expected to grow at an average of 4.39% every year reaching close to 1000K Tonnes by the year 2025.

Growth of Outbound Freight Traffic



The outbound international freight traffic in expected to grow at an average of 4.75% every year reaching close to 1400K Tonnes by the year 2025.

Chapter 13: Areas to Focus

India’s population is expected to increase further over the forecast horizon, rising from 1.3 billion currently to almost 1.5 billion by the end of 2025, along with the expected continuation of economic development and growth in household incomes.

A strong and growing business environment results in employment opportunities, investment and trade which the aviation industry can both help to enhance and benefit from. The job trend in the aviation industry is expected to be highly productive not only for their airline employers but for the economies in which they are employed.

The projections are expected to have a huge impact in the Indian market and also globally. With the projections we can develop as vision for 2025 that focuses in the key areas that for a short term that will help the aviation industry in India grow more after 2025.

- Cargo Services
- The Gulf Corridor
- Human Resource Management
- Skill Enhancement
- Aviation Security
- Manufacturing

13.1 Cargo Services

The capability of handling cargo in India is very less at the moment which is only about 4 Million Tonnes (International and Domestic). The efficiency of handling cargo should be increased by atleast 2 Million tonnes.

With the downfall of premiere airlines like Jet Airway and Air India, the cargo services offered by other domestic airlines will not be able to meet the demand in the coming years. Key measures should be taken in this case to enhance the cargo services at the airport as well as by the domestic carriers.

Key Measures:

- Enhance the ground handling capabilities of cargo services at the airports.
- Include dedicated cargo terminals at airports.
- Focusing more on the Gulf corridor to connect long haul cargo services.
- Focusing airports in the west coast, since huge traffic flow from the Gulf, US, UK and Europe flows through the airports in the west coast.

13.2 Human Resource Management

Managing human resources at the airports is a huge task in the aviation industry. One of the key problems is to manage Ground Handling. Ground Handling is one of the critical tasks at the airports, which includes passenger handling, cargo handling, cleaning, baggage handling, catering, security checks, etc.

With a potential of huge traffic growth, the ground handling in India is also expected to grow at a huge pace. Ground handling staff contributes to portions of jobs generated by the aviation industry.

Current ground handling capacity of all the major airports in India is at its threshold. More workforce is required in ground handling to handle the future demand. This will also lead to an increase in job opportunities and greater revenue for the airports.

13.3 The Gulf Corridor

The increase in airspace congestion in the Middle East will have a huge impact for aviation in India as the Gulf countries contribute to more than 35% of total International Passenger Traffic to and from India. United Arab Emirates (UAE) contributes to 30% of total International Freight Traffic. With an estimated traffic of more than 35 Million Passengers and 500K Tonnes of Freight by 2025 the Gulf corridor becomes one of the leveraged corridors.

India must take serious to develop the aviation management with the corridor. As there will be an increase in demand the Domestic Carriers must handle the competition from the Foreign Carriers.

Emirates is one of the major competitors to Air India in the Gulf corridor in terms of market share in both International Passenger and Freight Traffic. With sluggish growth in the debt ridden Air India, it will become difficult for the Domestic Carriers to gain the market share in the Gulf corridor.

Key Measures:

- Designing new and efficient airways connecting the Gulf countries.
- Improving the Air Traffic Management in the corridor by installing more ATC Towers in the west coast of India for efficient management of the airspace.
- Efficient scheduling of inbound and outbound flights as seen by the pattern in passenger traffic flows during Q4 and Q1 of every year.
- The Domestic carriers should introduce more long haul flights to Europe, UK and America with a connecting flight from the Gulf. This will help in the improvement of services in the Gulf corridor and also connect to far places.

13.4 Skill Enhancement

With an estimate of huge traffic flow for the next few years there will be an increase in demand for the skills needed to handle the traffic. India will require a large number of skilled Pilots, Engineers, Air Traffic Controllers, Ground Handling Staff, Security Forces, etc.

For this kind of growth, the central and state governments should invest more in academia by establishing new skill centres for aviation and also upgrading the current facilities and building infrastructure to prepare skilled human resources.

Infrastructure

India has many number of reputed aviation education institutes such as National Flying Training Institute (NFTI), Rajiv Gandhi National Aviation University (RGNAU). India has 30+ flying academies and other engineering and cabin crew training institutes.

The technology used in these aviation institutes should be upgraded to latest software and instruments. More number of flight simulators should be installed in these institutes to train the pilots to more level of accuracy.

Training Quality

Modernizing the infrastructure alone will not be enough to enhance the skill, the quality of training should also be improved. The training materials should be constantly updated with the update in technology and should include new methodologies for training.

The Government of India should help setup a research committee that can devise more effective training methods and evaluation metrics for the skills required.

Financial Aid

The Government of India should collaborate with the aviation education institutes to setup funds and help with the financial aid of students willing to enhance their skills in this field. Doing so will attract more number of students within India and also International students. This will help the institutes to increase its value and also establish India as one of the global training hubs of aviation.

13.5 Aviation Security

With a huge traffic flow it becomes very important for the airports to maintain the security and keep everything safe. With the increase in traffic there will an increase in passenger and staff movements at the airports which sometimes becomes difficult to control. There are steps that can be taken to efficiently manage the movements at the airport, thus maintaining security.

Security Check Mechanisms

Today there are a lot of evolving technologies with applications of Artificial Intelligence (AI). This technology can be used as an advantage in managing the security at the airports. Applications like Facial Recognition Software can be installed at the entry of the airports for faster and more effective way of security checks.

Automated Passenger Movement

There should more number of self-check-in counters for the passengers to check-in for their flights and get the boarding passes, rather than waiting in the check-in line. Automated movement of passengers will help ease the congestion at the airport thus making it easy for the security personnel to monitor movements within the airport.

Live Data Sharing

Live data of people and aircraft movements in the airport can be shared with the authorities who can constantly monitor them for any security threats or emergencies. This can be a lot of help to the emergency services in improving the response time.

13.6 Manufacturing

With global passenger and freight traffic expected to grow more than 6.3% and 4.6% every year respectively, the need for more commercial and cargo aircraft fleet will be required to meet the demand. As suggested in the analysis that the domestic carriers should invest more in the purchase of more cargo aircrafts, there will a huge potential for growth in the aviation manufacturing industry.

Financial Planning

With the estimated revenue to reach upto Rs15,000 Cr by 2025, the AAI should collaborate with the government to introduce policies that would help with the ease of doing business for the manufacturing industries. More allocation of the budget can be used to improve the production capabilities of parts in India.

Make In India

Make In India is a great initiative that could lead to a huge growth in India in terms of economy and GDP. Focusing more on the aviation industry in the Make In India scheme. This idea will establish India as a global hub for manufacturing. This would bring in more innovation into designing and manufacturing of aircrafts in India.

There is a huge potential that this could attract more sourcing and FDI (Foreign Direct Investment) for aviation start-ups in India.

Enhancing Capabilities of HAL

Hindustan Aeronautics Limited in the state run company, but its capability is limited to manufacturing only military technology and not civil aviation. The GOI should take steps in the next 5 year to enhance the capabilities of HAL in civilian aircraft manufacturing.

Chapter 14: Model Limitations

The time series data for the international passenger and freight traffic is available only from year 2014-2018 in the official data source. Due to the less availability of data, the long term forecast capability of the model is limited. Also there are a lot of external factors that might affect the demand.

There is come cyclic pattern observed in the data than is not captured by the model. Since these patterns are dynamic and unpredictable and happens due to some external economic and business conditions.

Chapter 15: Factors Influencing Future Demand

There are some factors that exist in the market and in general that influences the forecast and might change the demand. Since these factors are very dynamic in nature and also depends on the present market situation, it is very difficult to capture those in building a forecast model.

Based on the current market situation and economic conditions some precautions can be taken to handle the estimated demand. Some of the factors that may influence the demand are:

- Air Services agreement
- Trade agreement
- Tourism policies
- Economic growth
- GDP
- Travel industry development
- Development strategies of airlines
- Infrastructure developments
- Bilateral relations with other countries

Chapter 16: Enhancements in Future Forecast Models

The present forecast model completely builds on the historical traffic data, i.e. the model uses only the past values to predict the future values. The model does not use any exogenous variables.

The present model can be further improved by adding some exogenous variable that influence the traffic demand. One of the best variable that has a very high correlation with our target variable is the GDP. GDP is expected to have a hypothetically very high correlation with the passenger and freight traffic demand. Hence the future model can include these features to improve the forecast accuracy.

Chapter 17: Conclusion

The analysis shows clear and high growth rates of the Aviation Industry in terms of passenger and freight demand for the next few years Industry despite being one of the most complicated industries. The potential of growth it has is very huge and also its contribution to the economic in huge.

The analysis shows how India should be prepared to handle the future demand by bringing in some strategies and changes in the management of the aviation industry. It supports and enhances business and investment decisions and also promotes the transfer of knowledge and innovation, and provides opportunities consistent growth of India.

However this demand will also bring in a lot of challenges for the aviation industry in terms of handling traffic at airports, for airlines to invest more in purchasing new aircrafts and for the manufacturing industry to constantly design and develop advance technologies for more efficiency.

The analysis focuses on how the policy makers in the government should play an important for enhancing the capabilities of the industry to meet the demand. The policies should not affect the growth and development and reduce the benefits that the aviation industry can deliver.

With the right policies and focus on the aviation industry, India can one of the global hubs of aviation and establish itself as a premier institution of the aviation industry leading to a high economic growth in the country.

Chapter 18: References

- Director General of Civil Aviation
- IBEF Aviation Presentation www.ibef.org
- https://school.geekwall.in/p/rJK_T1qZQ/time-series-prediction-with-arima-models-in-python
- Transportation Research E-Circular, August 2002
- India’s Air Transport Sector, IATA, January 2016
- International Transport Forum , OECD
- Air Navigation Case Study, ICAD, January 2013
- https://www.iata.org/en/pressroom/pr/2019-09-05-01
- Time Series Tutorial https://courses.analyticsvidhya.com/courses/creating-time-series-forecast-using-python