ANALYSIS AND PREDICTION OF ROAD ACCIDENTS IN INDIA – A TIME SERIES STUDY

Report submitted to

SISTER NIVEDITA UNIVERSITY

For the award of the degree

of

Master of Science in Statistics

by

Asad Amaan

Student Id. : 2321209002002

Registration No. : 230020023876

Under the supervision of

Prof. Aniruddha Choudhuri

DEPARTMENT OF STATISTICS

SISTER NIVEDITA UNIVERSITY

NEWTOWN, KOLKATA, WEST BENGAL

A Satyam Roychowdhury Initiative

SNU
SISTER NIVEDITA UNIVERSITY

M.Sc. FINAL YEAR MASTER PROJECT/ DISSERTATION COMPLETION CERTIFICATE

This is to certify that **ASAD AMAAN** (**Registration No: 230020023876**) has prepared the Master Project work entitled **Analysis and Prediction of Road Accidents in India – A Time Series Study** under the supervision of **Prof. Aniruddha Choudhuri** based on the survey of literatures in his/her area of interest for the partial fulfillment of the M.Sc. Degree in Statistics from Sister Nivedita University.

AGhosal 29/11/24

Dr. Anindita Ghosal

Assistant Professor and HOD

M. Pal 29/11/2024

Signature of External with date

29/11/24

Signature of Supervisor with date

Acknowledgement

I would like to express my heartfelt gratitude to all those who contributed to the successful completion of this project on **Analysis and Prediction of Road Accidents in India – A Time Series Study**.

First and foremost, I would like to thank my supervisor, **Prof. Aniruddha Choudhuri**, for his invaluable guidance, encouragement and constructive feedback throughout the course of this project. His expertise and insights were instrumental in shaping the direction and quality of this work.

I extend my sincere thanks to **Sister Nivedita University** for providing the resources and platform needed for conducting this study.

A special note of appreciation goes to my family, friends and peers namely Procheto Nag, Nikhat Azad, Nafis Ahmed and Anika Kalim who supported me and motivated me during this process. Their encouragement helped me stay focused and committed to achieving my goals.

This project has been a valuable learning experience, and I am deeply grateful for the opportunity to contribute to the understanding of road safety issues in India.

Asad Amaan

M.Sc. Statistics Sem-III,

Department of Statistics,

Sister Nivedita University

Table of Contents

Abstract

Road accidents have become very common now-a-days. As more and more people are buying automobiles, the incidences of road accidents are increasing day-by-day. Furthermore, people have also become more careless now and there are a lot of people who does not like to follow the traffic rules. Especially, in big cities, there are various modes of transports making the roads look narrower. Thus, road accidents are bound to happen; you pick up a newspaper, you will find at least one or two news about road accidents daily. They cause loss of life as well as material. People need to be more careful when on the road, no matter which mode of transport they are using. Even people who are walking, are not safe because of the rise in these incidents. Every day people witness accidents in the news, from their relatives and even with their own eyes. Keeping these things is mind, we thought of doing a project based on the road accidents occurring in all the states and union territories of India. In this project, we have taken the data on road accidents occurring in the 35 states and union territories of India during the years 2001-2014. Initially we got two types of datasets on the road accidents in India – one having the number of accidents occurring in a 3-hourly basis for every year for all the states and union territories and another having the same occurring in monthly basis for all the years for each state and union territories. For the representation part of the data, we here have drawn Multiple Line diagrams, Multiple Divided Bar diagrams and Pie diagrams and tried to draw out the features appearing from the said diagrams. The next part of this project deals with finding out a suitable time series model corresponding to the monthly data for the years 2001-2014 for each state and union territories separately and use that model to predict the amount of road accidents that would occur in the next three years 2015-2017 on a monthly basis. For the said purpose, we have considered the ARIMA models of time series analysis with suitably chosen parameters. This project can be helpful to common people in understanding the problems caused by road accidents and also can be thought as a way of awareness for general public of our country.

Introduction

Road accidents are a significant public health and safety concern, particularly in a vast and diverse country like India. Every year, thousands of lives are lost, and countless others are affected by injuries and economic losses resulting from traffic collisions. The complexity of road accident dynamics arises from a range of factors, including traffic density, road infrastructure, driver behavior, and environmental conditions, which vary widely across different states. Understanding and addressing these challenges requires a data-driven approach that can identify trends, predict future occurrences, and aid in the formulation of effective prevention strategies.

Time series analysis is a powerful statistical tool used to study data points collected over time. It enables the identification of patterns, trends, seasonality, and irregularities in data, making it particularly well-suited for analyzing historical road accident data. By leveraging this method, we can gain valuable insights into how accident rates have evolved over time and forecast future trends. This approach not only provides a clearer understanding of past and current accident scenarios but also equips policymakers, law enforcement, and urban planners with the ability to take proactive measures to enhance road safety.

In this study, we focus on analyzing historical road accident data from various states in India using time series analysis techniques. The objective is to uncover underlying patterns and develop predictive models that can forecast future road accidents. This analysis aims to answer critical questions such as: Which states show an increasing trend in road accidents? Are there specific times of the year when accidents are more likely to occur? How can predictive insights inform targeted interventions?

By bridging the gap between data analysis and actionable strategies, this study emphasizes the role of advanced statistical tools in addressing a pressing societal issue. The findings will not only provide a deeper understanding of the state-wise dynamics of road accidents in India but also contribute to the broader goal of enhancing road safety and reducing the burden of accidents nationwide.

Objectives

India faces a pressing issue of road accidents, which are often attributed to a combination of factors such as poor road infrastructure, reckless driving, and inadequate enforcement of traffic rules. By analyzing historical data on road accidents, we can identify underlying patterns and trends, enabling us to forecast future occurrences and implement targeted interventions.

Research Objectives

- Exploratory Data Analysis (EDA): To gain insights into the historical trends, seasonal variations, and other patterns in road accident data.
- Time Series Modeling: To develop appropriate time series models, such as ARIMA, to capture the temporal dependencies in the data.
- Model Evaluation: To assess the accuracy and predictive power of the selected models using suitable performance metrics.
- Future Predictions: To forecast future trends in road accidents and identify potential hotspots or periods of increased risk.

Data Description

Here we have considered two types of data sets on road accidents for 35 states and union territories of India over the years 2001 to 2014.

The 1st data set contains information on road accidents corresponding to all the 35 states and union territories of India over the years 2001 to 2014 in 3-hourly basis (0-3 hrs., 3-6 hrs. and so on)

The 2nd data set contains information on road accidents corresponding to all the 35 states and union territories of India over the years 2001 to 2014 in monthly basis (January, February and so on)

The variables involved in the above mentioned two data sets are enumerated below:

STATE/UT : Name of State or Union Territory for which the data of road accident is given. Total of 35 States and Union Territory are mention in the data

YEAR : The year for which the data of road accident is given. From the year 2001 to2014.

0-3 hrs. (Night) : The data of road accident collected from 0 hour to 3 hour midnight.

3-6 hrs. (Night) : The data of road accident collected from 3 hour to 6 hour in the morning.

6-9 hrs (Day) : The data of road accident collected from 6 hour to 9 hour in the morning.

9-12 hrs (Day) : The data of road accident collected from 9 hour to 12 hour noon.

12-15 hrs (Day) : The data of road accident collected from 12 hour to 15 hour in the evening.

15-18 hrs (Day) : The data of road accident collected from 15 hour to 18 hour in the evening.

18-21 hrs (Night) : The data of road accident collected from 18 hour to 21 hour at night.

21-24 hrs (Night) : The data of road accident collected from 21 hour to 24 hour at night.

Total : total number of road accident in a particular State/UT in a particular year.

JANUARY :The data of road accidents collected on the month of January for the year 2001 to 2014.

FEBRUARY :The data of road accidents collected on the month of January for the year 2001 to 2014.

MARCH :The data of road accidents collected on the month of January for the year 2001 to 2014.

APRIL :The data of road accidents collected on the month of January for the year 2001 to 2014.

MAY :The data of road accidents collected on the month of January for the year 2001 to 2014.

JUNE :The data of road accidents collected on the month of January for the year 2001 to 2014.

JULY :The data of road accidents collected on the month of January for the year 2001 to 2014.

AUGUST :The data of road accidents collected on the month of January for the year 2001 to 2014.

SEPTEMBER :The data of road accidents collected on the month of January for the year 2001 to 2014.

OCTOBER :The data of road accidents collected on the month of January for the year 2001 to 2014.

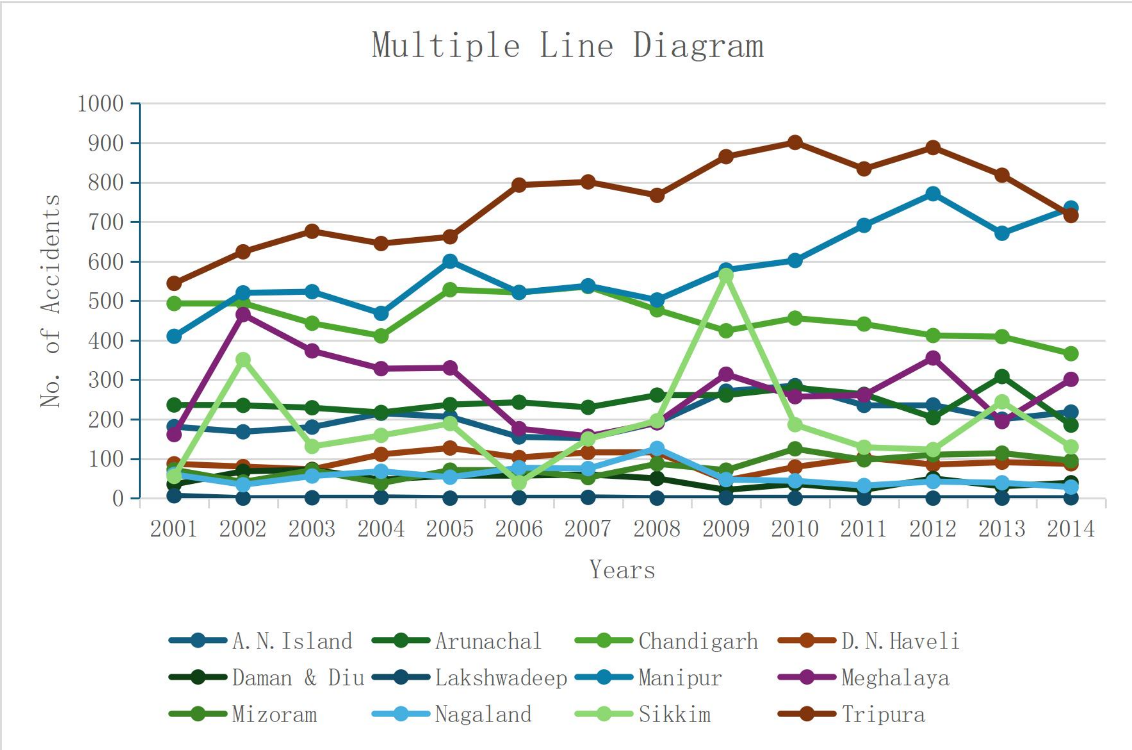
NOVEMBER :The data of road accidents collected on the month of January for the year 2001 to 2014.

DECEMBER :The data of road accidents collected on the month of January for the year 2001 to 2014.

Data Source: *https://www.kaggle.com/datasets/manugupta/road-accidents-in-india*

Graphical Representation of data

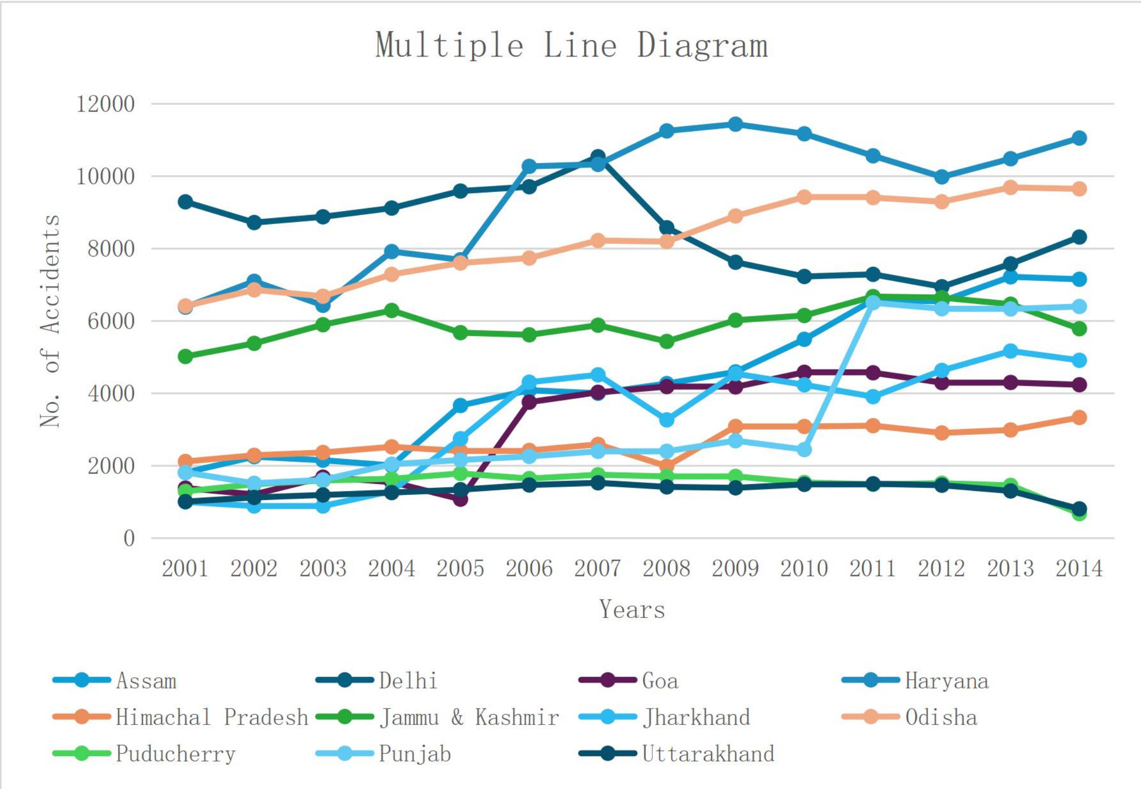
Diagram: Representing the states having minimum road accidents



In twelve (12) States/UTs the total number of road accidents every year from the year 2001 to 2014 l lies between 0 to 1000. Among which Lakshwadeep had minimum no. of accidents and Tripura had maximum no. of accidents.

From the diagram we can also see that there is sudden increase in the total number of accidents in the year 2002 and 2009 for Sikkim.

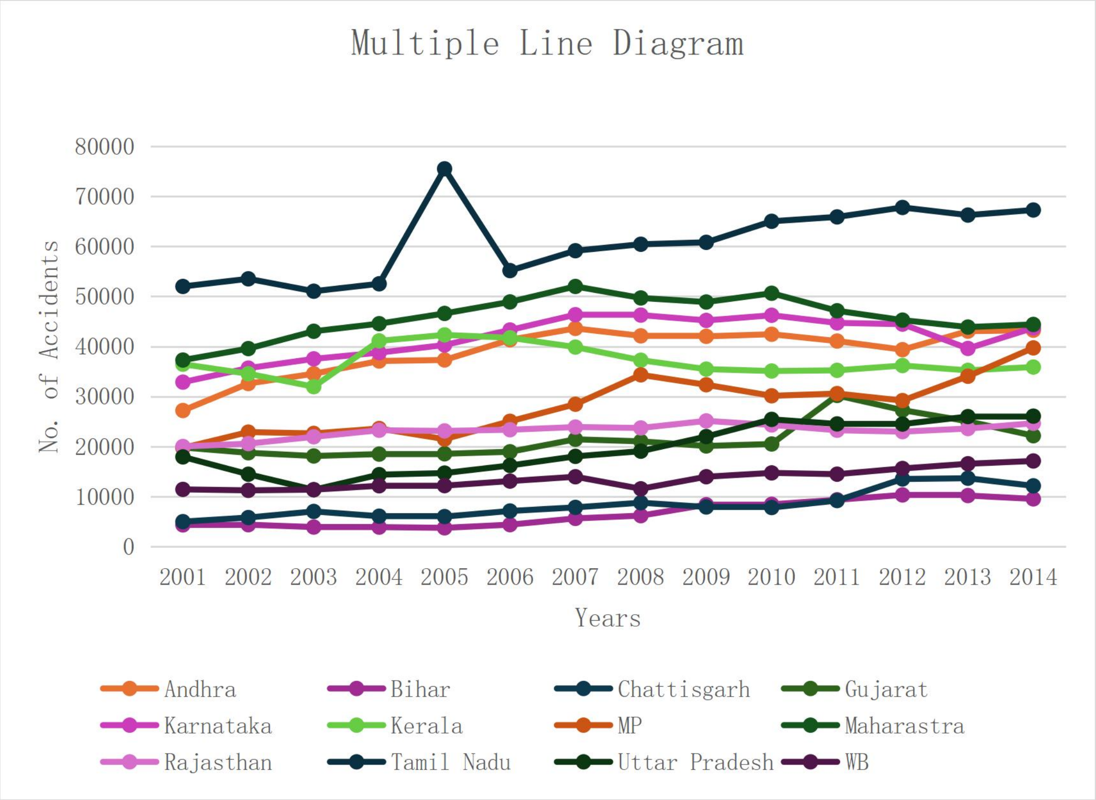
Diagram: Representing the states having average road accidents



In eleven (11) States/UTs the total number of road accidents every year from the year 2001 to 2014 lies between 1000 to 12000. Among which Uttarakhand had minimum no. of accidents and Haryana had maximum no. of accidents.

From the diagram we can also see that there is sudden increase in the total number of accidents in the year 2011 for Punjab.

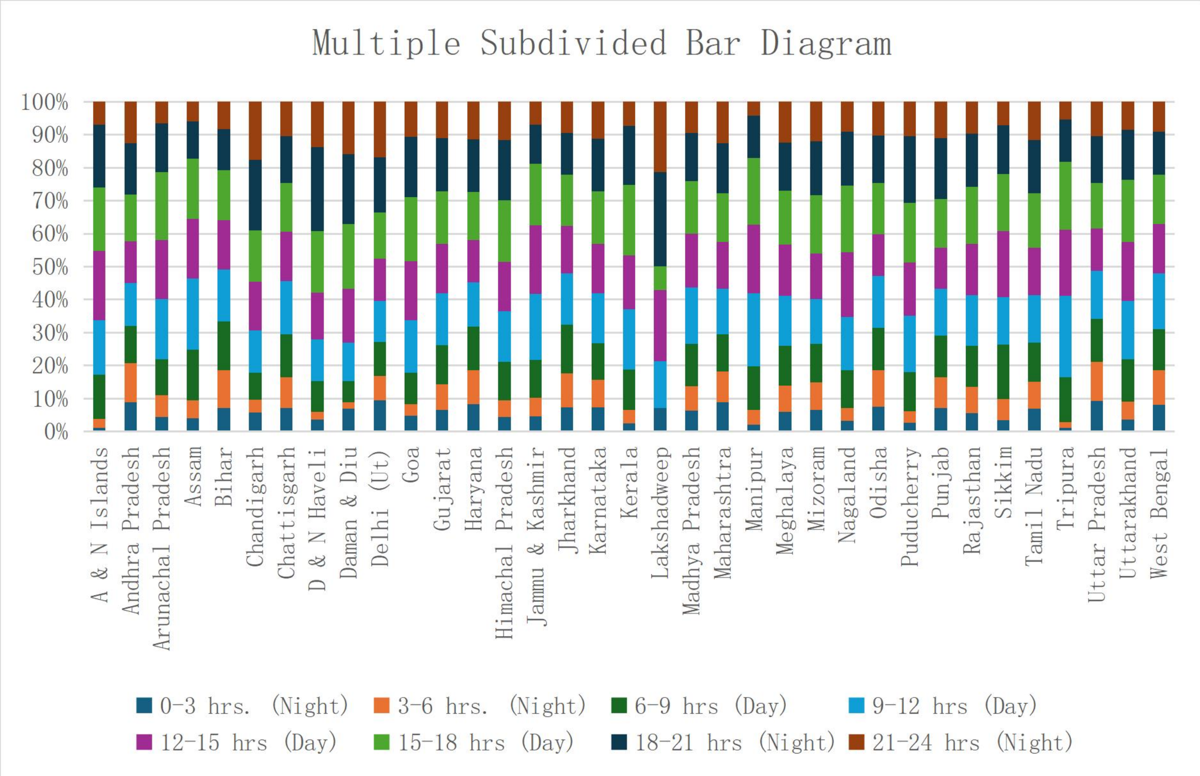
Diagram: Representing the states having maximum road accidents



This multiple line diagram contains the States/UTs having maximum total number of road accidents every year from the year 2001 to 2014, among which Chattisgarh had minimum no. of accidents and Tamil Nadu had maximum no. of accidents.

From the diagram we can also see that there is sudden increase in the total number of accidents in the year 2005 for Tamil Nadu.

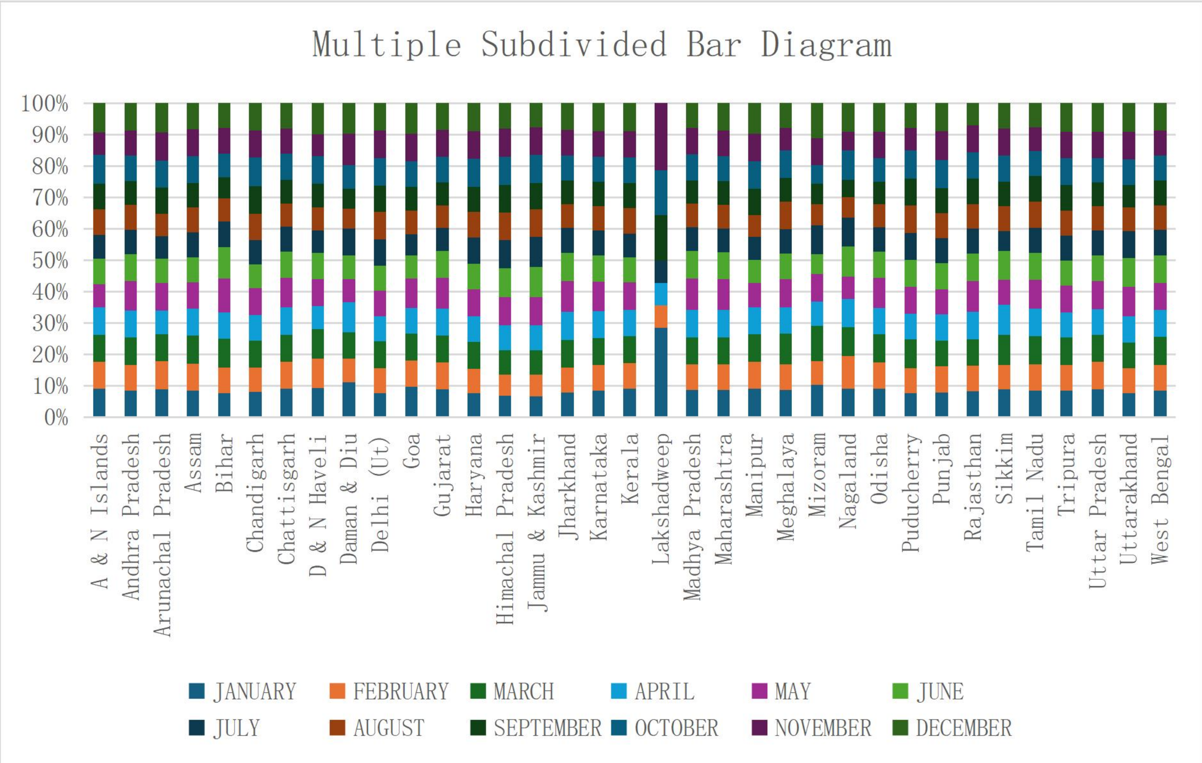
Hourly Subdivided Bar Diagram of Road Accidents for 35 States and UT of India



From the subdivided bar diagram we can observe that for Tamil Nadu the maximum number of accidents occur between 15-18 hrs. and for West Bengal the maximum number of accidents occur between 9-12 hrs.

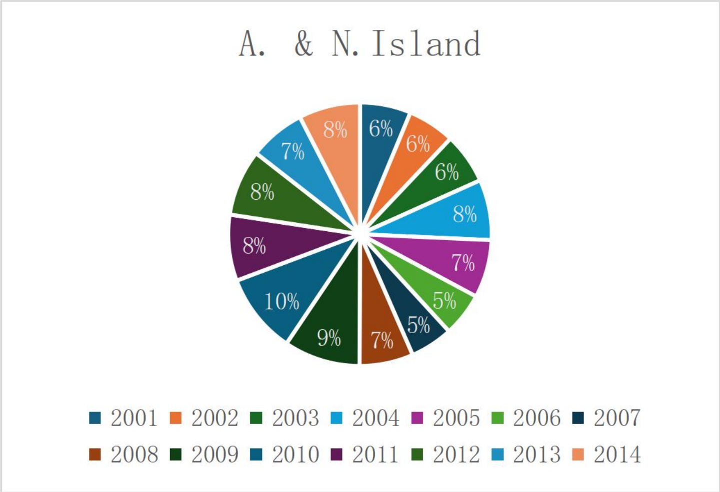
Diagram No. 5:

Monthly Subdivided Bar Diagram of Road Accidents for 35 States and UT of India

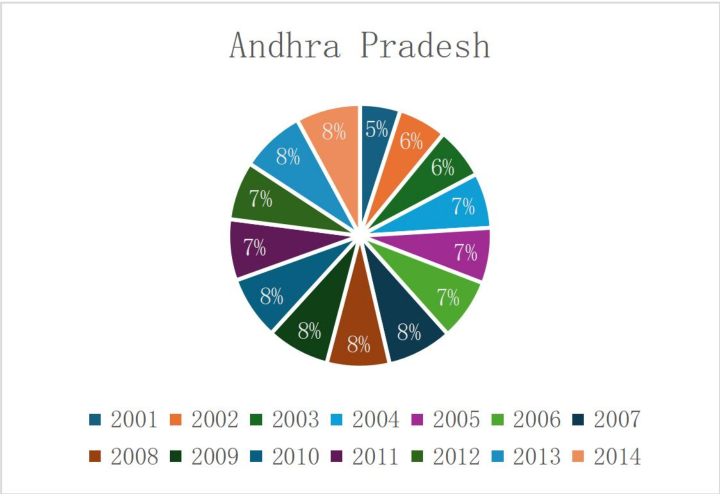


From the subdivided bar diagram we can observe that for Tamil Nadu the maximum number of accidents occur in the month of April and for West Bengal the maximum number of accidents occur in the month of February.

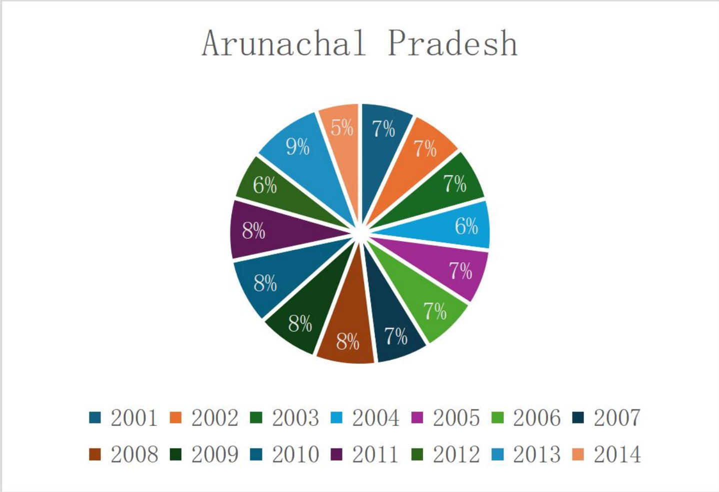
Pie Diagrams for each States and UTs.



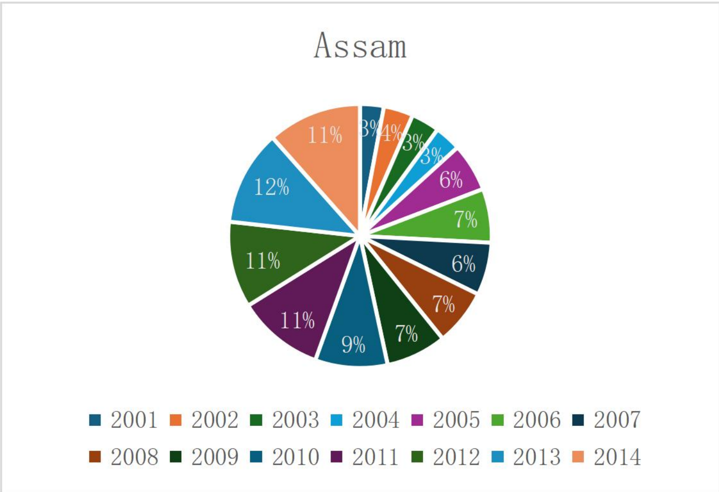
- From the pie chart of **Andaman and Nicobar Island,** we can see that, 8% ,9% ,10% ,8% ,8% ,8% accidents occur during the year 2004 ,2009 ,2010 ,2011 ,2012 ,2014 respectively. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2005 ,2006 ,2007 ,2008 ,2013.



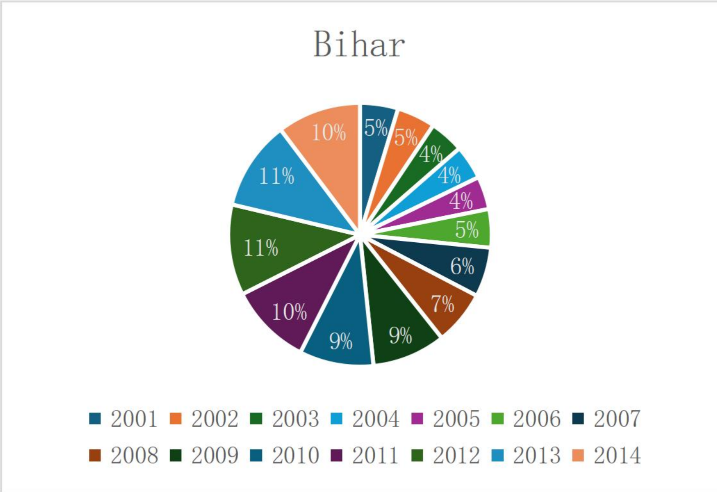
- From the pie chart of **Andhra Pradesh,** we can see that, 8% accidents occur during each of the year 2007 ,2008 ,2009 ,2010 ,2013 ,2014 . The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2006 ,2011 ,2012 .



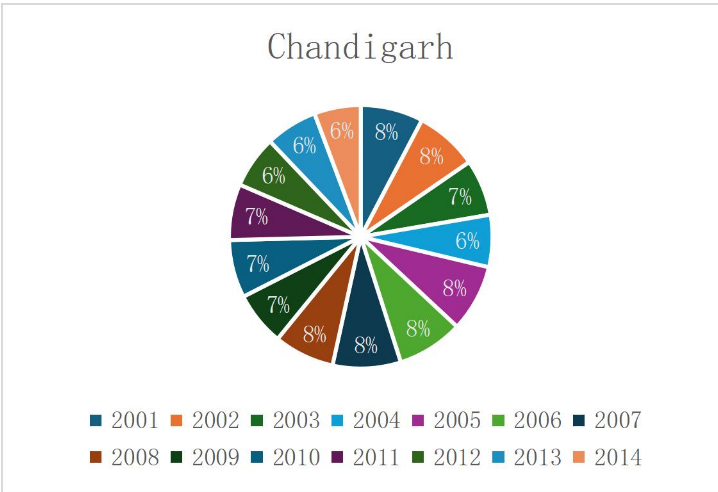
- From the pie chart of **Arunachal Pradesh,** we can see that, 8% ,8% ,8% ,8% ,9% accidents occur during the year 2008 ,2009 ,2010 ,2011 ,2013 respectively. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2006 ,2007 ,2012 ,2014.



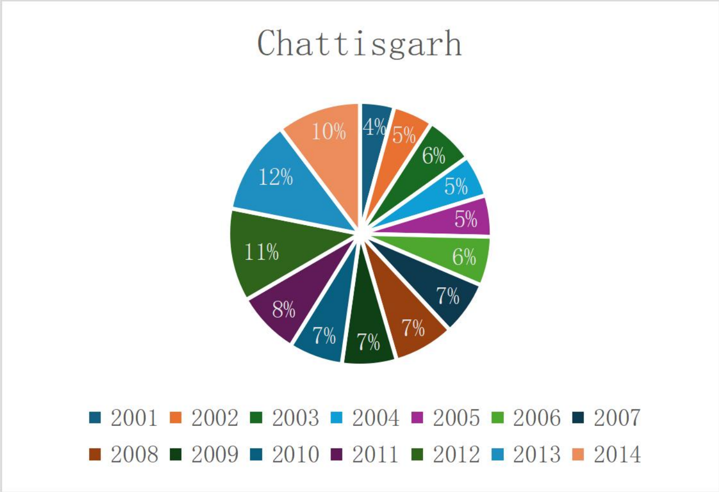
- From the pie chart of **Assam,** we can see that, 11% ,11% ,12% ,11% accidents occur during the year 2011 ,2012 ,2013 ,2014 respectively. The road accidents occur below 10% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2006 ,2007 ,2008 ,2009 ,2010.



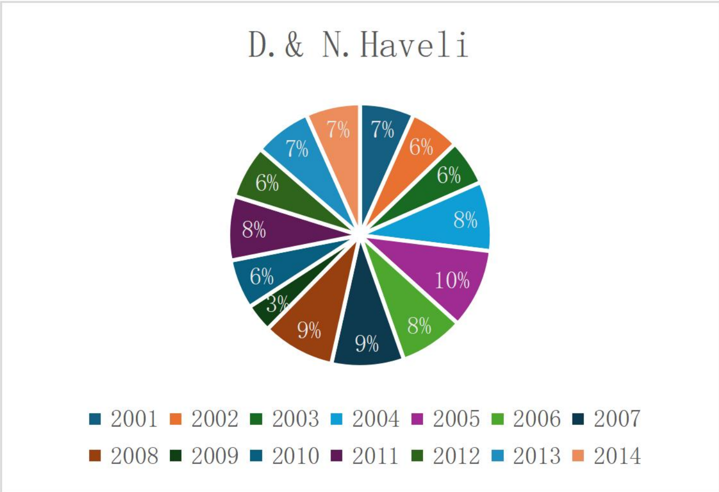
- From the pie chart of **<u>Bihar,</u>** we can see that, 10% ,11% ,11% ,10% accidents occur during the year 2011 ,2012 ,2013 ,2014 respectively. The road accidents occur below 10% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2006 ,2007 ,2008 ,2009 ,2010.



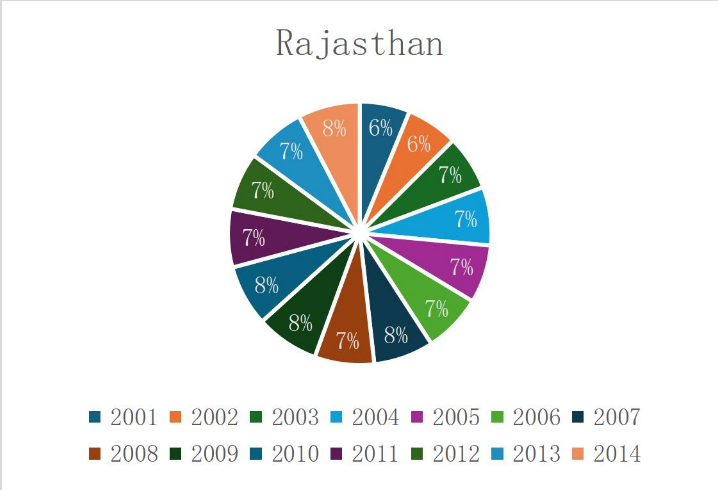
- From the pie chart of **<u>Chandigarh,</u>** we can see that, 8%accidents occur during each of the year 2001 ,2002 ,2005 ,2006 ,2007 ,2008 respectively. The road accidents occur below 8% during the year 2003 ,2004 ,2008 ,2009 ,2010 ,2011 ,2012 ,2013 ,2014.



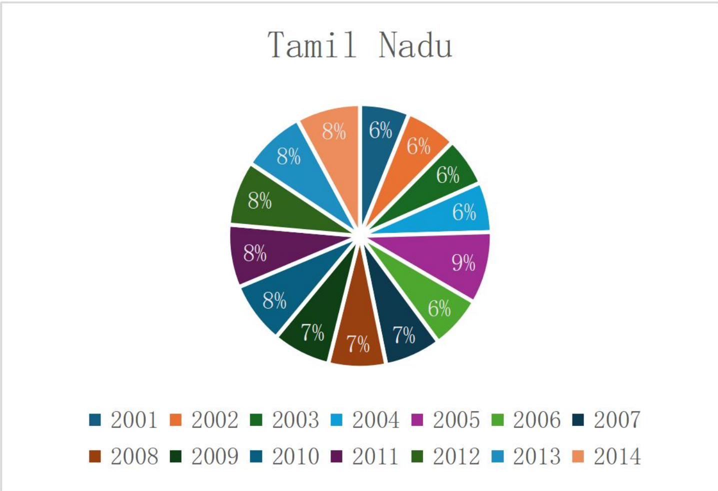
- From the pie chart of **Chattisgarh,** we can see that, 8% ,11% ,12% ,10% accidents occur during the year 2011 ,2012 ,2013 ,2014 respectively. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2006 ,2007 ,2008 ,2009 ,2010 .



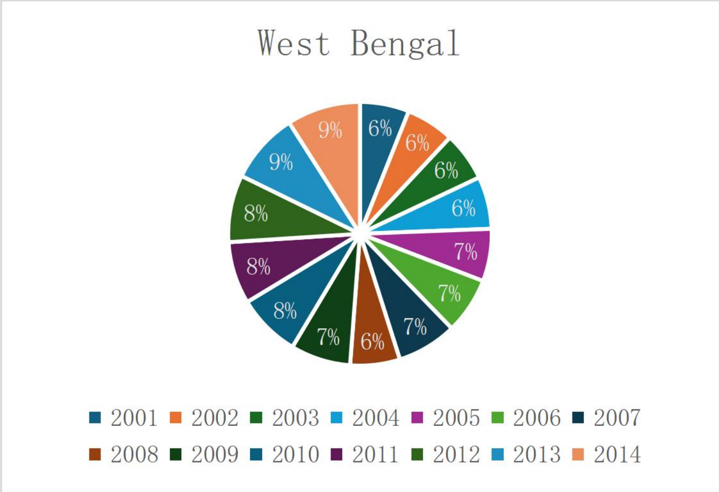
- From the pie chart of **Dadar and Nagar Haveli,** we can see that, 8% , 10% ,8% ,9% ,9% ,8% accidents occur during the year 2004 ,2005 ,2006 ,2007 ,2008 ,2011 respectively. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2009 ,2010 ,2012 ,2013 ,2014.



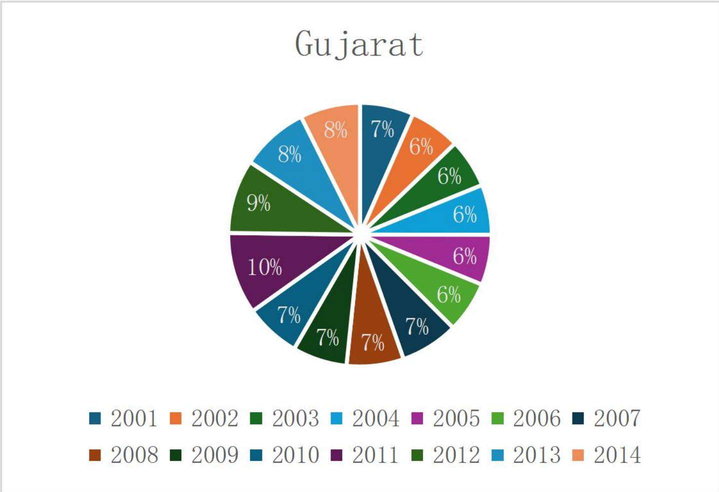
- From the pie chart of **Rajasthan,** we can see that, the total number of road accidents over the years 2001 to 2014 is almost constant.



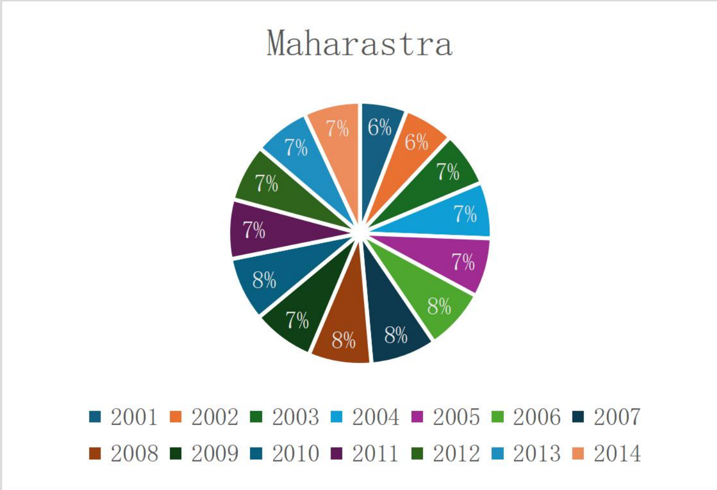
- From the pie chart of **Tamil Nadu,** we can see that, 8% accidents occur during each of the year 2010 ,2011 ,2012 ,2013 ,2014 respectively and 9% in the year 2005. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2004 ,2006 ,2007 ,2008 ,2009 .



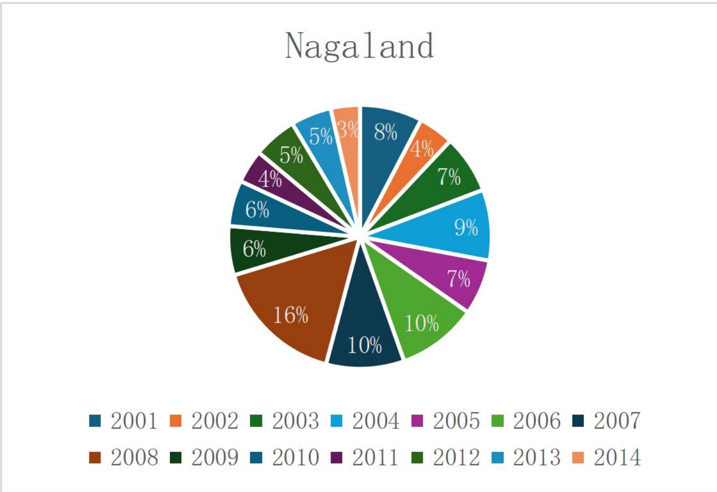
- From the pie chart of **West Bengal,** we can see that, 8% ,8% ,8% ,9% ,9% accidents occur during the year 2010 ,2011 ,2012 ,2013 ,2014 respectively. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2006 ,2007 ,2008 ,2009.



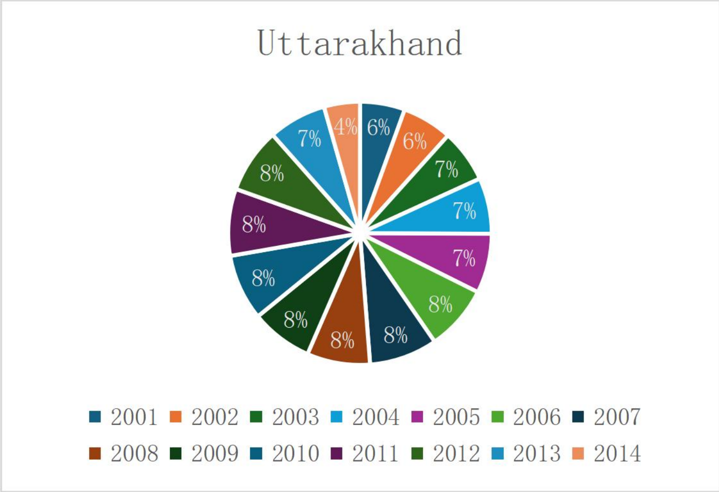
- From the pie chart of **Gujarat,** we can see that,10% ,9% ,8% ,8% accidents occur during the year 2011 ,2012 ,2013 ,2014 respectively. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2006 ,2007 ,2008 ,2009 ,2010.



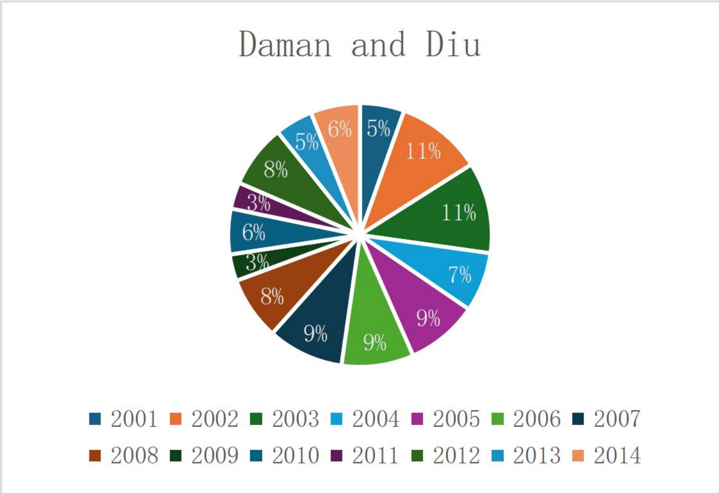
- From the pie chart of **Maharashtra,** we can see that, 8% ,accidents occur during each of the the year 2006 ,2007 ,2008 ,2011 respectively. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2009 ,2010 ,2012 ,2013 ,2014.



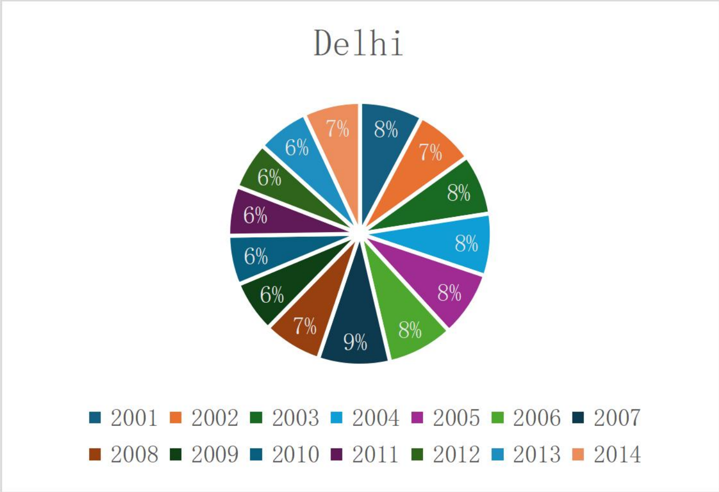
- From the pie chart of **Nagaland,** we can see that, 8% ,9% ,10% ,10% accidents occur during the year 2001 ,2004 ,2006 ,2007 respectively. In the year 2008 the total number of road accidents goes up to 16%. The road accidents occur below 8% during the year 2002 ,2003 ,2003 ,2009 ,2010 ,2011 ,2012 ,2013 ,2014.



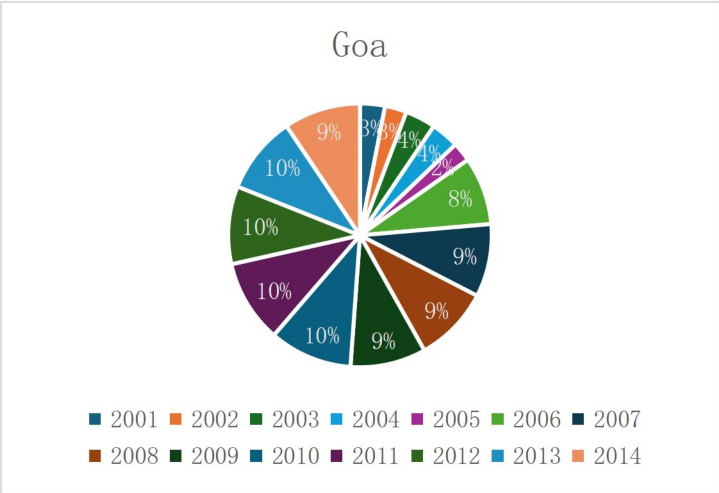
- From the pie chart of **Uttarakhand,** we can see that, 6% ,6% ,4% accidents occur during the year 2001 ,2002 ,2014 respectively. The road accidents occur at the rate of 7% and 8% from the year 2003 to2013 .



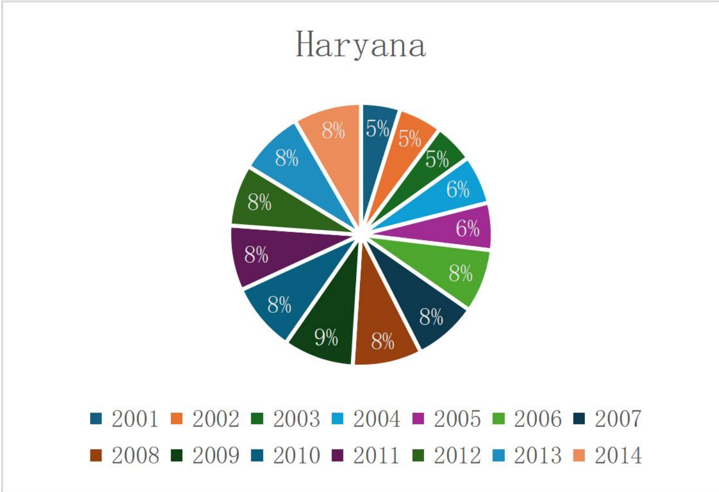
- From the pie chart of **Daman and Diu,** we can see that, 5%, 7%, 9%, 9%, 9%, 8%, 6%, 8%, 5%, 6% accidents occur during the years 2001, 2004, 2005, 2006, 2007, 2008, 2010, 2012, 2013, 2014 respectively. We can see in the years 2009 and 2011 the rate of accidents goes down to 3% and in the years 2002 and 2003 the road accidents goes up to 11%



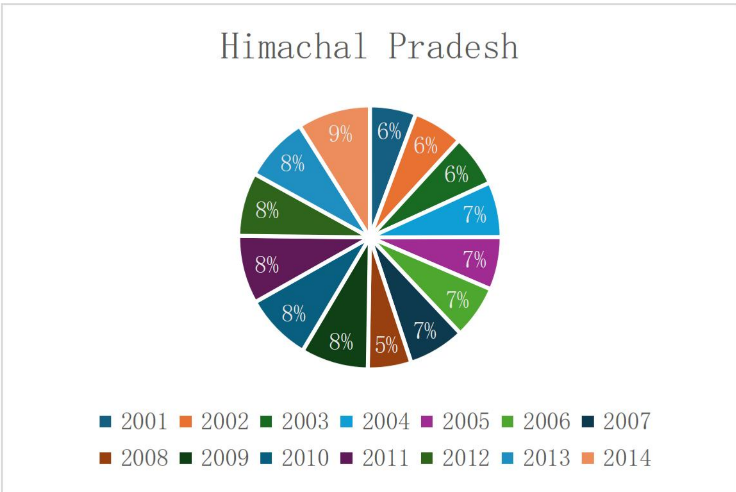
- From the pie chart of **Delhi,** we can see that, 8% accidents occur during each of the year 2001 ,2003 ,2004 ,2005 ,2006 ,2014 and 9% in the year 2007. The road accidents occur below 8% during the year 2002 ,2008 ,2009 ,2010 ,2011 ,2012 ,2013 ,2014.



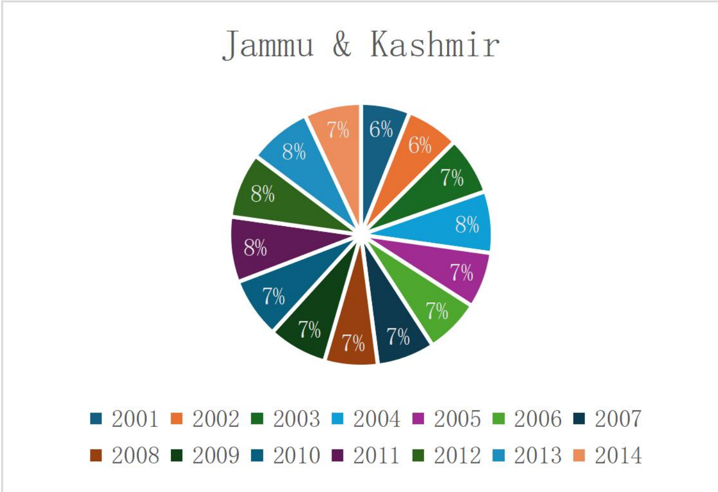
- From the pie chart of **Goa,** we can see that, 10% accidents occur during each of the year 2010 ,2011 ,2012 ,2013 respectively. The road accidents occur below 10% during the year 2001 to 2009 and 2014.



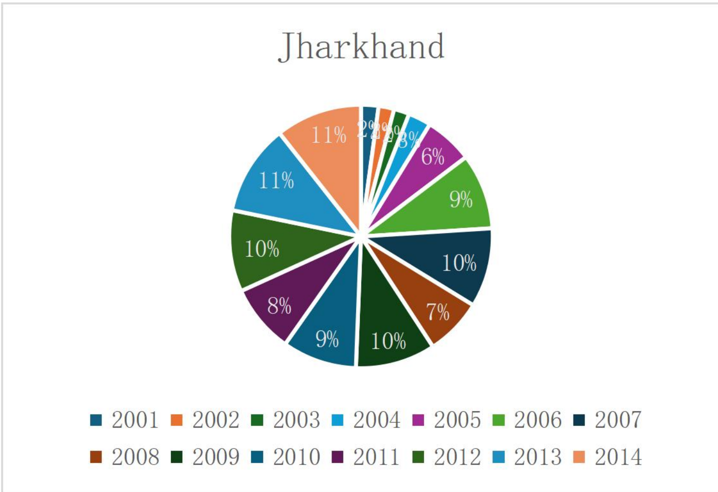
- From the pie chart of **Haryana,** we can see that, 5% and 6% accidents occur in the year 2001 to 2005 respectively. The road accidents occur above 7% during the year 2006 to 2014.



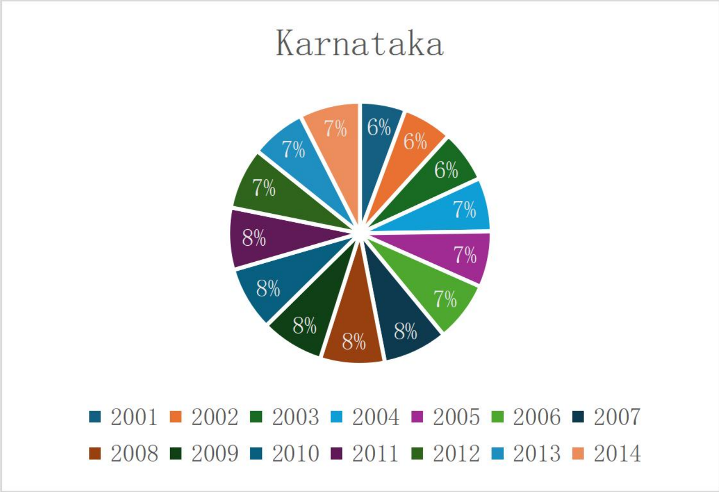
- From the pie chart of **Himachal Pradesh,** we can see that, 8% accidents occur during the year 2009 to 2013 and 9% in the year 2014 . The road accidents occur below 8% from the year 2001 to 2008 .



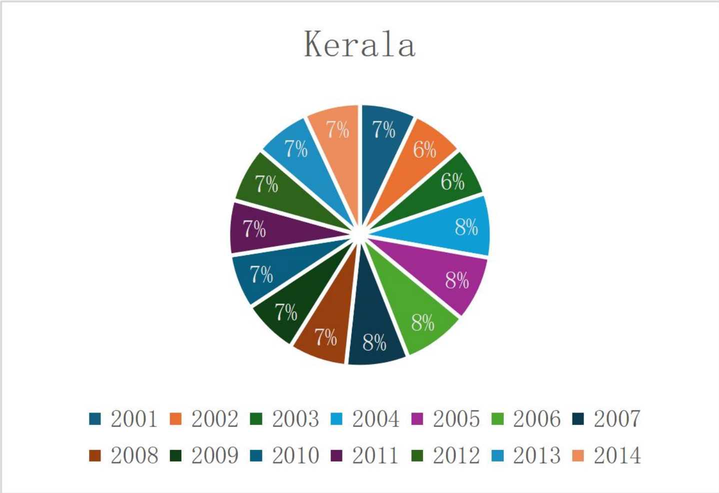
- From the pie chart of **Jammu and Kashmir,** we can see that, the total number of road accidents over the years 2001 to 2014 is almost constant.



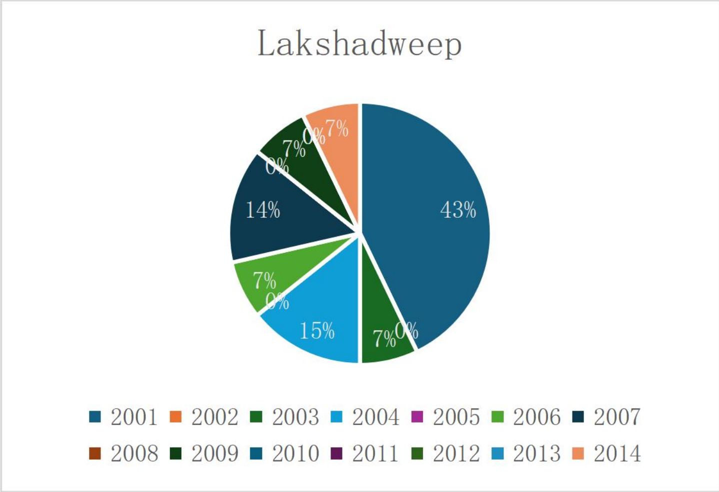
- From the pie chart of **Jharkhand,** we can see that, 10% ,10% ,10% ,11% ,11% accidents occur during the year 2007 ,2009 ,2012 ,2013 ,2014 respectively. The road accidents occur below 10% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2006 ,2008 ,2010 ,2011. In the year 2001 , 2002 and 2003 the rate of accidents is very less.



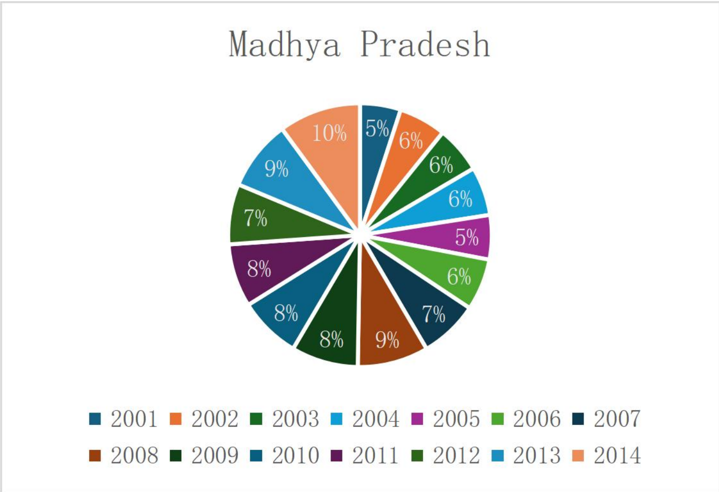
- From the pie chart of **<u>Karnataka,</u>** we can see that, 8% accidents occur in the year 2007 to 2011. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2004 ,2005 ,2006 ,2012 ,2013 ,2014.



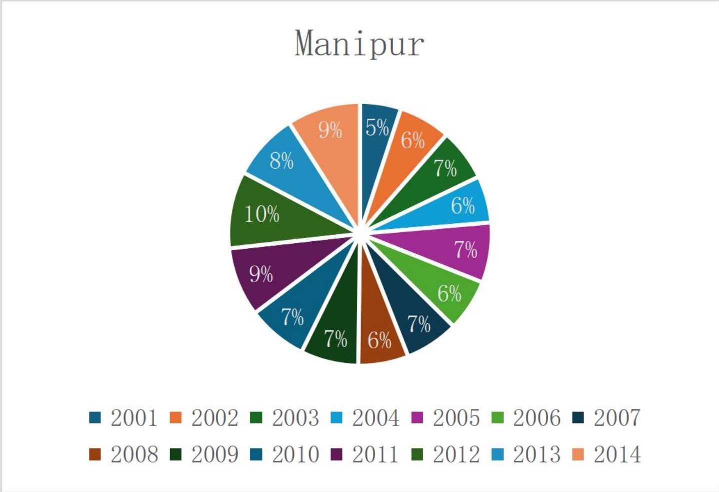
- From the pie chart of **<u>Kerela,</u>** we can see that, 8% accidents occur in the year 2004 to 2007. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2008 ,2009 ,2010 ,2011 ,2012 ,2013 ,2014.



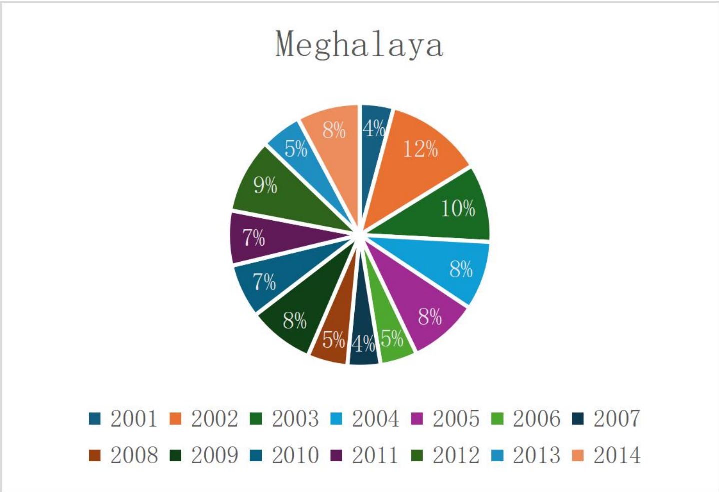
- From the pie chart of Lakshadweep**,** we can see that, maximum number of accidents occur in the year 2001. In the following years the rate went down to 7% 15% ,7% , 14% ,7% 7% in the year 2003 ,2004 ,2006 ,2007 ,2012 ,2014.



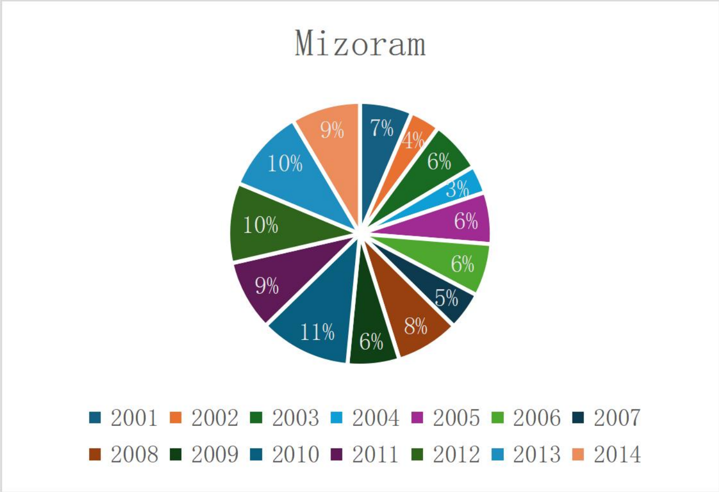
- From the pie chart of **Madhya Pradesh,** we can see that, 8% accidents occur during the year 2009 ,2010 ,2011. The road accidents occur below 8% during the year 2001 ,2002 ,2003 ,2005 ,2006 ,2007, 2012 . During the years 2008, 2013 and 2014 the number of accidents went above 8%.



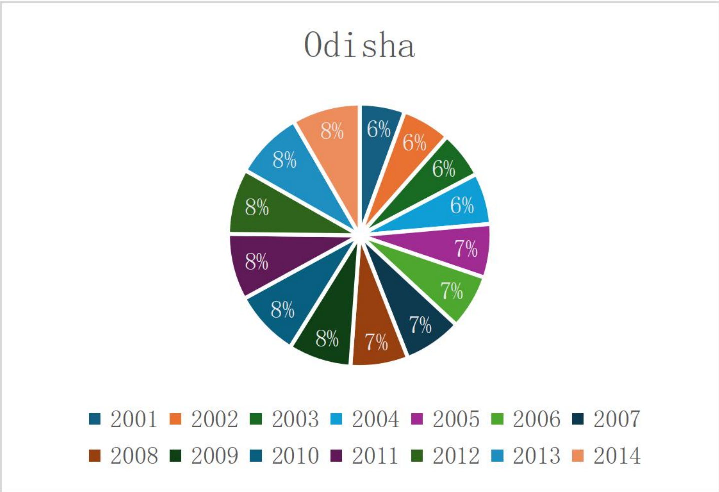
- From the pie chart of **Manipur,** we can see that the rate of road accidents in the year 2001, 2002, 2003, 2004, 2005, 2006,2 007, 2008, 2009 ,2010 are below 8%. The road accidents in the years 2011, 2012, 2013, 2014 are 8% or above.



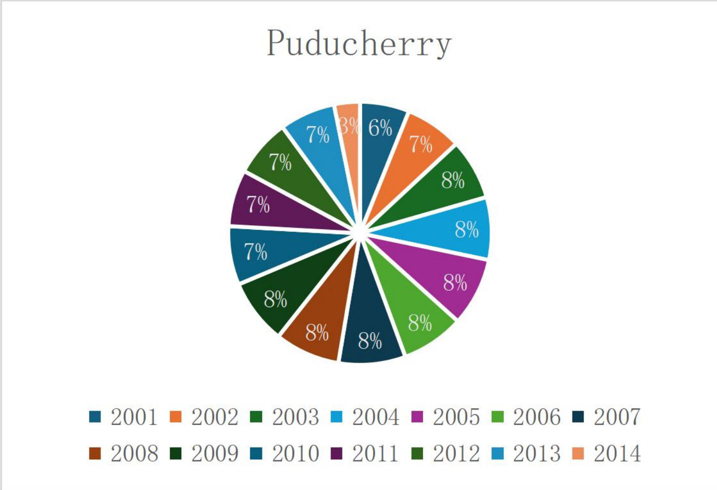
- From the pie chart of **Meghalaya,** we can see that, 12% , 10%, 8%, 8%, 8%, 9% ,8% accidents occur during the years 2002, 2003, 2004, 2005, 2009, 2012, 2014 respectively. The road accidents occur below 8% during the years 2001 ,2006, 2007, 2008, 2010,2011,2013.



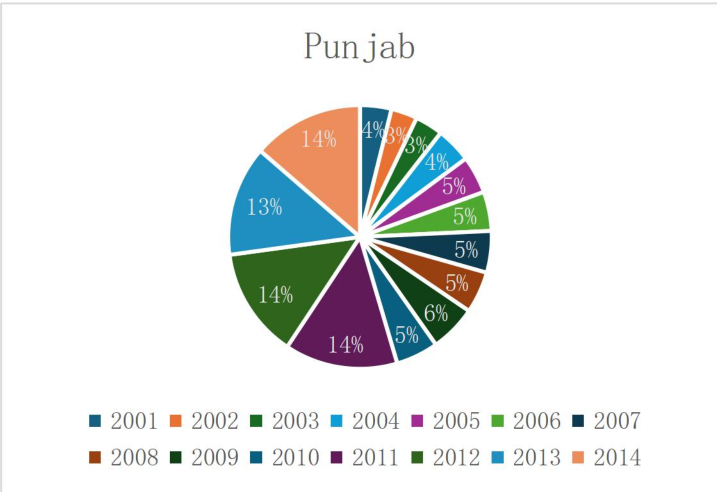
- From the pie chart of **<u>Mizoram,</u>** we can see that, above 8% accidents occur during the years 2010 to 2014. The road accidents occur below 9% during the year 2001 to 2009.



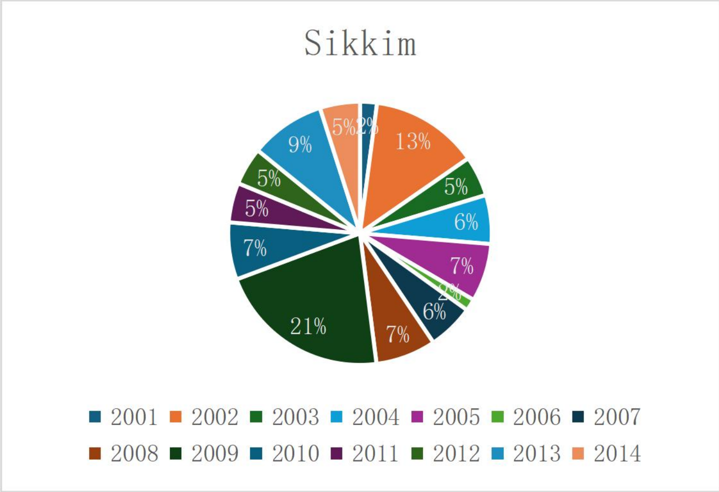
- From the pie chart of **<u>Odisha,</u>** we can see that, the total number of road accidents over the years 2001 to 2014 is almost constant.



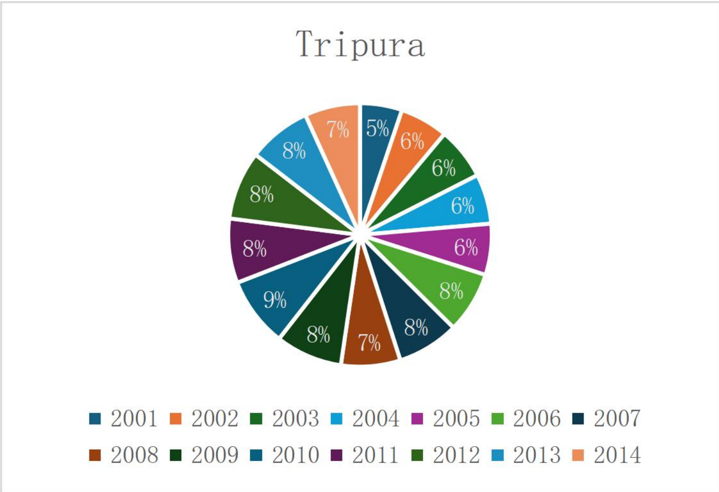
- From the pie chart of **Puducherry,** we can see that, the total number of road accidents over the years 2001 to 2014 is almost constant.



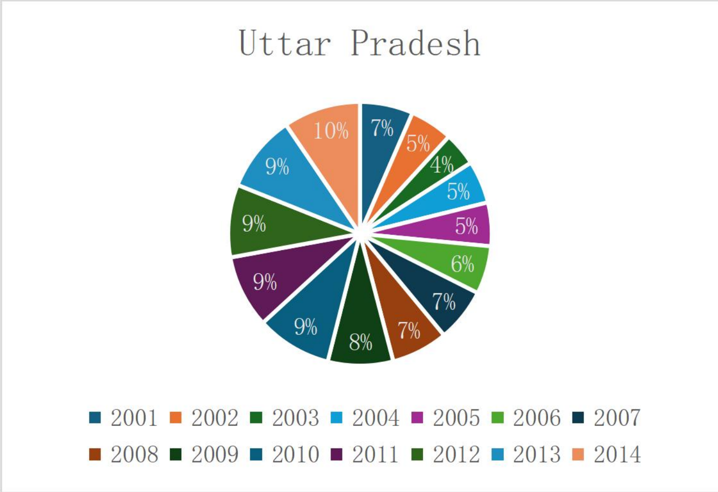
- From the pie chart of **Punjab,** we can see that, above 10% accidents occur during the years 2011, 2012, 2013, 2014. The road accidents occur below 8% during the years 2001, 2002, 2003, 2004, 2005, 2006, 2007,2008, 2009, 2010.



- From the pie chart of **Sikkim,** we can see that, 13%, 21% accidents occur during the year 2002, 2009 respectively. The road accidents occur below 10% during the years 2001, 2003, 2004, 2005, 2006, 2007, 2008, 2010, 2011 ,2012, 2013, 2014.



- From the pie chart of **Tripura,** we can see that, below 9% road accidents occur during the years 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2011, 2012, 2013, 2014. The road accidents occur 9% in the year 2010.



- From the pie chart of **Uttar Pradesh,** we can see that, 9% accidents occur during the year 2010 ,2011 ,2012, 2013 and 10% in the year 2014. The road accidents occur below 9% during the years 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009.

Modelling of Time Series Data

- ## Finding an appropriate model & prediction for road accidents:

Here we want to predict road accidents for each state and UTS. So, we will take the monthly data for 2001 to 2014 of each state and UTs to find a reasonable model to forecast. We can fit ARIMA model for each state and UTs and then we will predict using this model.

Auto-regressive Integrated Moving Average (ARIMA) Prediction Model:

✧ What is an Auto-regressive Integrated Moving Average (ARIMA):

An auto-regressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends.

A statistical model is auto-regressive if it predicts future values based on past values. For example, an ARIMA model might seek to predict a stock's future prices based on its past performance or forecast a company's earnings based on past periods.

✧ Understanding Auto-regressive Integrated Moving Average

An auto-regressive integrated moving average model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.

An ARIMA model can be understood by outlining each of its components as follows

- Auto-regression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
- Integrated (1): represents the differencing of raw observations to allow the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
- Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

✧ ARIMA parameters:

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

- p: the number of lag observations in the model, also known as the lag order.
- d: the number of times the raw observations are differenced; also known as the degree of differencing.
- q: the size of the moving average window, also known as the order of the moving average.

For example, a linear regression model includes the number and type of terms. A value of zero (0), which can be used as a parameter, would mean that particular component should not be used in the model. This way, the ARIMA model can be constructed to perform the function of an ARMA model, or even simple AR, I, or MA models.

✧ ARIMA and Stationary Data:

In an auto-regressive integrated moving average model, the data are differenced in order to make it stationary. A model that shows stationarity is one that shows there is constancy to the data over time. Most economic and market data show trends, so the purpose of differencing is to remove any trends or seasonal structures.

Seasonality, or when data show regular and predictable patterns that repeat over a calendar year, could negatively affect the regression model. If a trend appears and stationarity is not evident, many of the computations throughout the process cannot be made and produce the intended results.

✧ What is ARIMA used for?

ARIMA is a method for forecasting or predicting future outcomes based on a historical time series. It is based on the statistical concept of serial correlation, where past data points influence future data points.

✧ How to build an ARIMA model & predict:

We will use Box-Jenkins model procedure to fit ARIMA model

- Box-Jenkins Model:

For building the Box-Jenkins Model, we will go through some steps.

Those are,

1. Identification: At first, we will check the seasonality of the data by visualizing the plot for data and a.c.f and p.a.c.f plot for the data. If seasonality is present then we will remove it by decompose method from the data.

Further we will proceed with the new data (seasonality free data) and we will plot a.c.f & p.a.c.f plot. Then we will check if the new data is stationary or not by visualization and unit root test (Augmented Dickey Fuller Test). If new data is not stationary then make stationary by taking difference. And then we will decide the order of the ARIMA model by checking a.c.f & p.a.c.f plot and the difference (from difference we will decide for 'd', from a.c.f plot we will decide 'q' & from p.a.c.f we will decide for 'p').

- ◆ Note: If there is no seasonality in original data, then we will proceed with original data.

2. Estimation: We will estimate the unknown coefficients to find the model.

3. Diagnostic checking: We will take the residual to check if the model is good or not.

If the model is a good fit, then a.c.f plot of residual should have a single spike at h=0 and for h>0 it should be insignificant. We will also check this by plotting a.c.f of residuals and by testing Ljung-Box Test. If we will get the p value of the test >0.05 then we can say that model is good, if not go back to step 1 and identify an alternative model. We will also choose the model with smallest AIC.

- Ljung-Box Test Statistic: For some m, $Q_{LB} = n(n+2)\sum_{h=1}^{m}(\rho(h))^2/(n-h)$, $\rho(h)$ is the a.c.f. function of residuals. (there are n residuals)

After fitting a model by this method we can forecast for future,

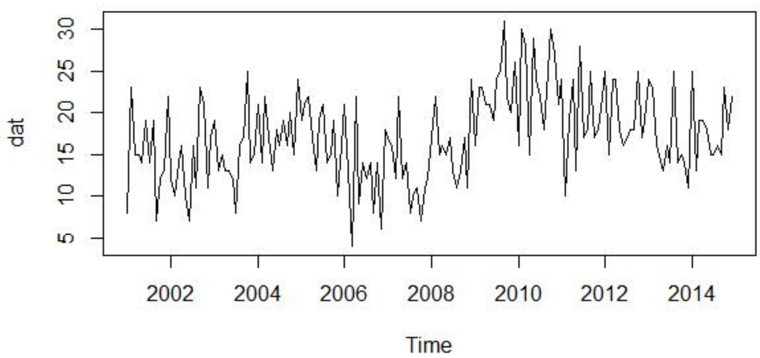
- Forecast: We will take the predicted values from the fitted ARIMA model. But it will give us seasonality free predictions, i.e., as we have deleted the seasonality part at first and have worked with the new data. So, after getting predictions from the model we will add the seasonality to get real predictions of our original data.
    - Note: If we have used the original data to fit the ARIMA model, i.e., there was no seasonality in the original data, then we will take just the prediction values from the fitted model.

Now we are going to fit ARIMA model for 35 states and predict road accidents for the next 3 years (2015-2017).

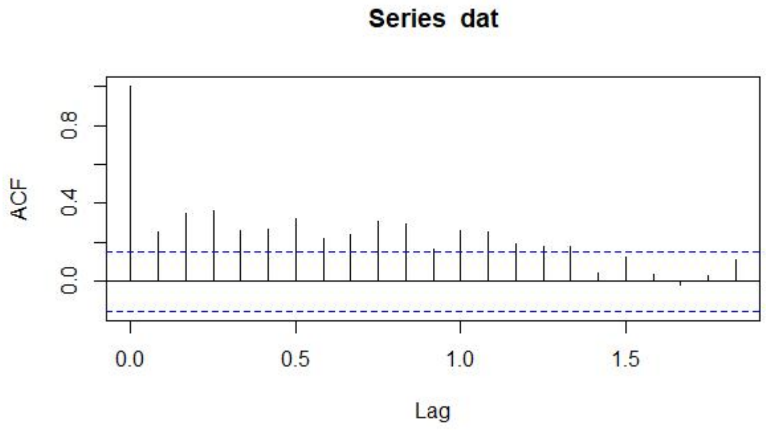
- **Andaman and Nicobar Islands :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

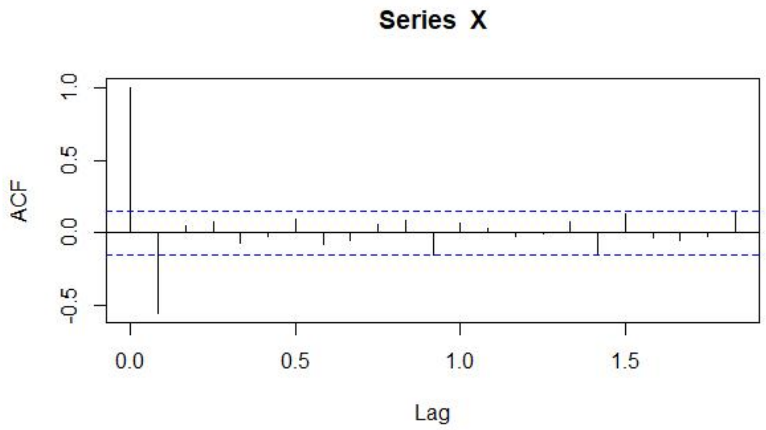
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differencing the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

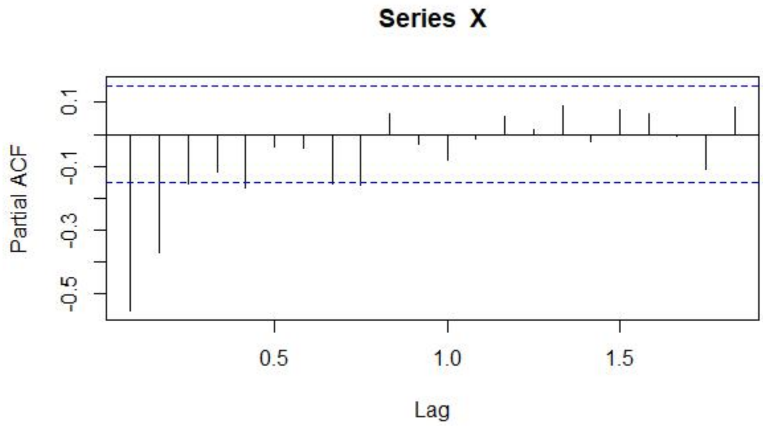
Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -2.926, Lag order = 5, p-value = 0.1898

alternative hypothesis: stationary

The pacf plot for the differenced data:



Now, we will take the diff to make it stationary.

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -7.7964, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary.

Therefore the data is stationary at d=1.

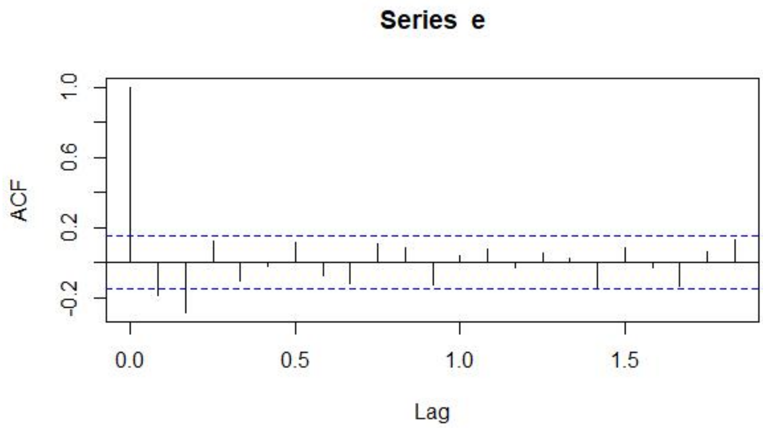
Fitted model by checking AIC is:

ARIMA(X,order = c(1,1,1)), so here we have taken p=1, d=1, q=1

Coefficients:

	ar1	ma1
	-0.1522	-0.7872
s.e.	0.0918	0.0568

sigma^2 estimated as 22.14: log likelihood = -496.21, aic = 998.42For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.021184, df = 1, p-value = 0.8843

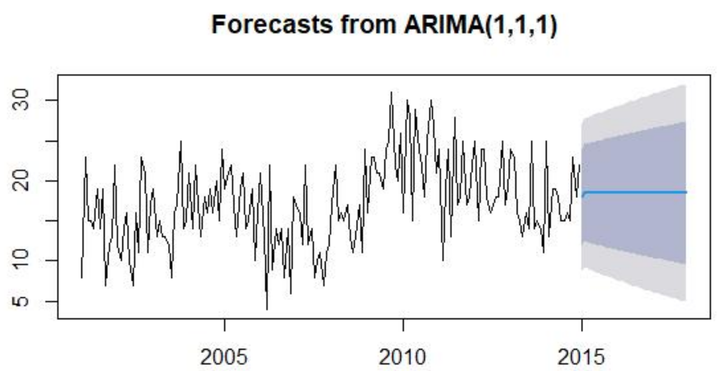
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	17.93767	18.55598	18.46187	18.47620	18.47402	18.47435	18.47430	18.47430
2016	18.47430	18.47430	18.47430	18.47430	18.47430	18.47430	18.47430	18.47430
2017	18.47430	18.47430	18.47430	18.47430	18.47430	18.47430	18.47430	18.47430

	Sep	Oct	Nov	Dec
2015	18.47430	18.47430	18.47430	18.47430
2016	18.47430	18.47430	18.47430	18.47430
2017	18.47430	18.47430	18.47430	18.47430

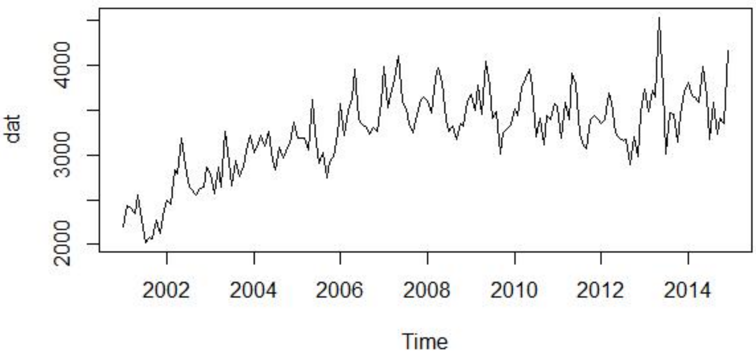
Plot for forecast is given below:



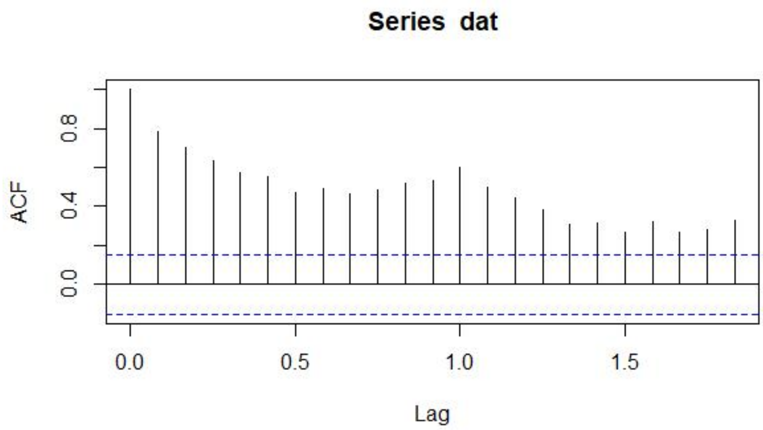
- **Andhra Pradesh:**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

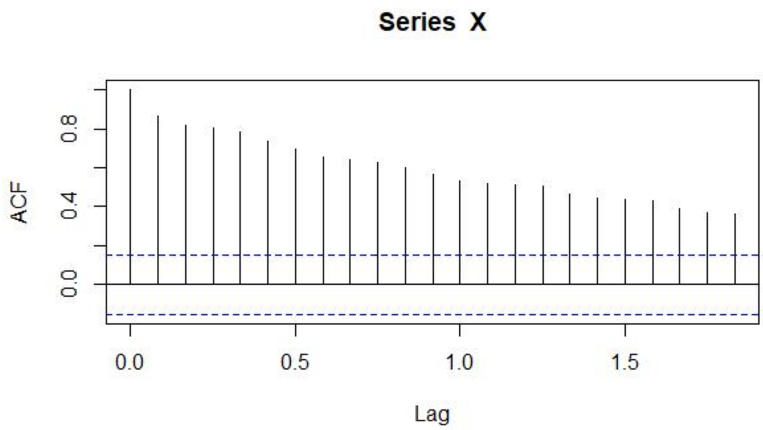


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.



So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -2.7421, Lag order = 5, p-value = 0.2665

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

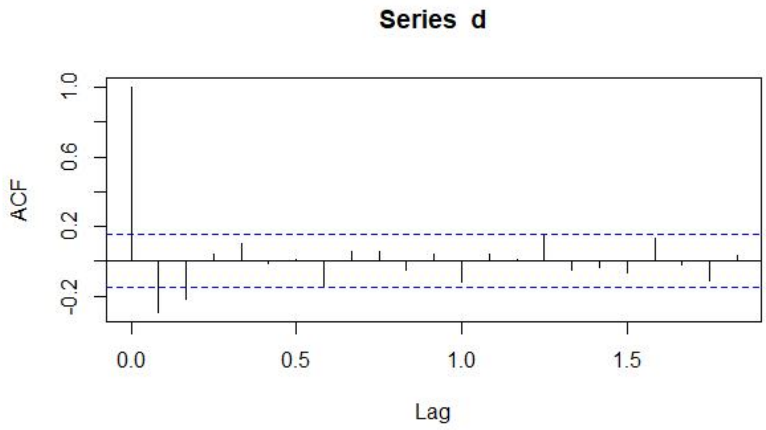
Augmented Dickey-Fuller Test

data: d

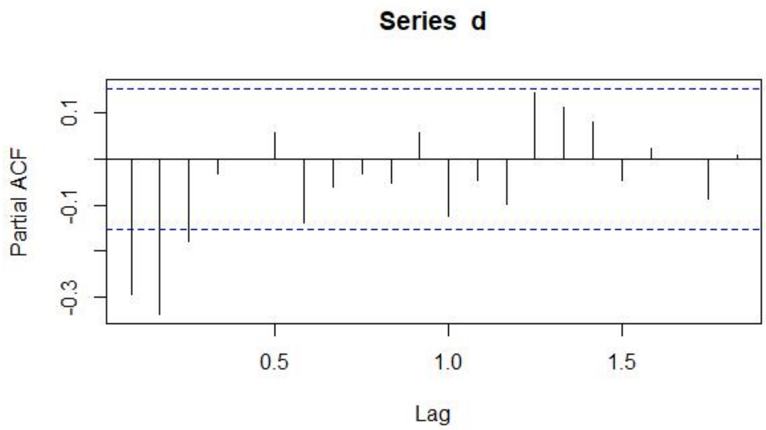
Dickey-Fuller = -5.8788, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary.

After differencing the data we again take the acf plot.



The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

ARIMA(x = dat, order = c(1, 1, 1))

Coefficients:

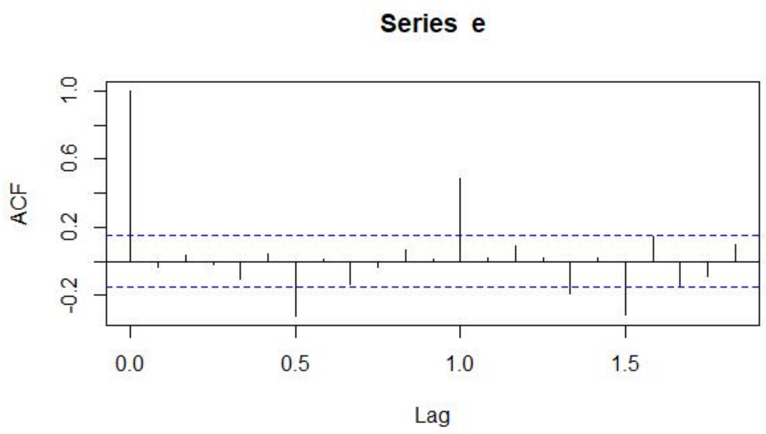
ar1 ma1

0.4504 -0.8763

s.e. 0.0871 0.0393

sigma^2 estimated as 63726: log likelihood = -1161.01, aic = 2328.02

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.22592, df = 1, p-value = 0.6346

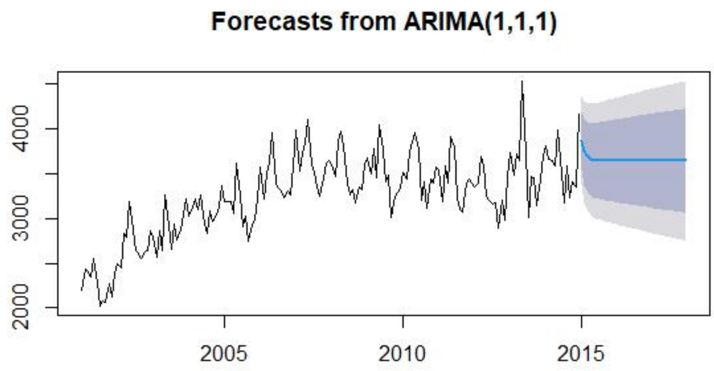
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3872.635	3744.113	3686.229	3660.160	3648.419	3643.131	3640.749	3639.676
2016	3638.814	3638.805	3638.801	3638.799	3638.798	3638.798	3638.798	3638.798
2017	3638.798	3638.798	3638.798	3638.798	3638.798	3638.798	3638.798	3638.798

	Sep	Oct	Nov	Dec
2015	3639.193	3638.976	3638.878	3638.834
2016	3638.798	3638.798	3638.798	3638.798
2017	3638.798	3638.798	3638.798	3638.798

Note that, this plot is for deseasonalized data:



Now, the original predictions are, i.e. after adding the seasonal part is given below,

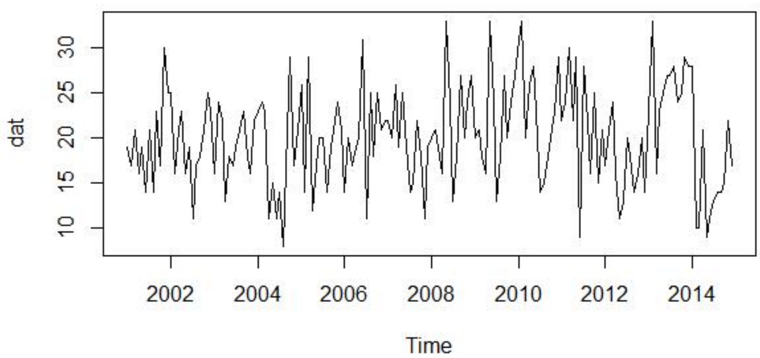
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3991.806	3698.784	3871.359	3777.568	4092.884	3723.689	3422.641	3501.206
2016	3757.985	3593.476	3823.930	3756.207	4083.264	3719.357	3420.690	3500.328
2017	3757.969	3593.469	3823.927	3756.206	4083.263	3719.356	3420.690	3500.327

	Sep	Oct	Nov	Dec
2015	3334.005	3479.112	3499.696	3698.242
2016	3333.610	3478.933	3499.616	3698.206
2017	3333.610	3478.933	3499.616	3698.206

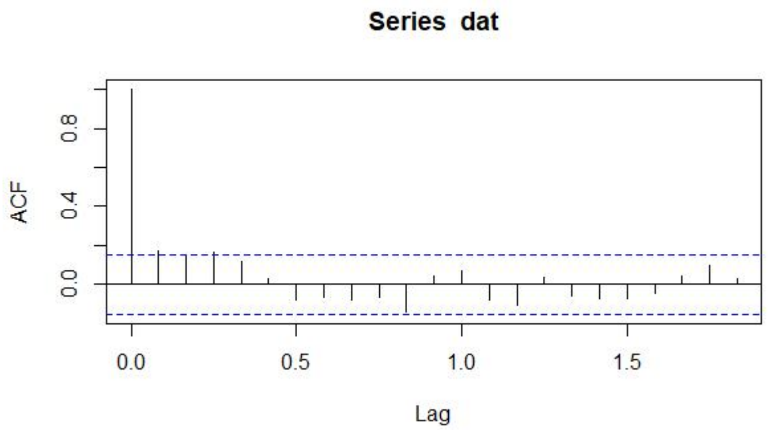
➢ **<u>Arunachal Pradesh:</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

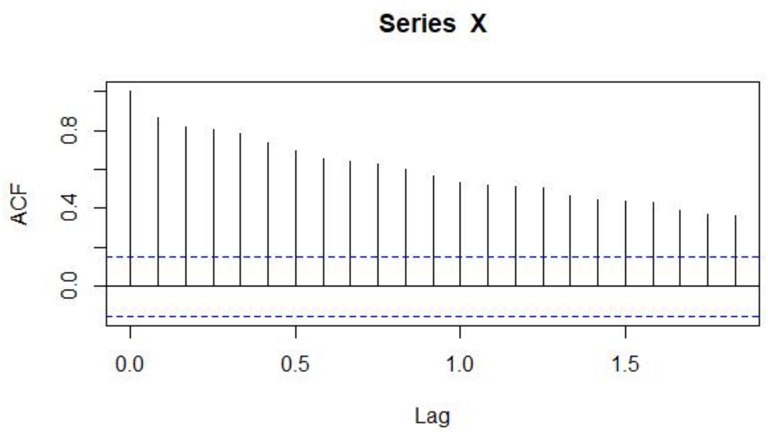


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.



So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



So, now we can say that the data is stationary, augmented Dickey Fuller test also showed that

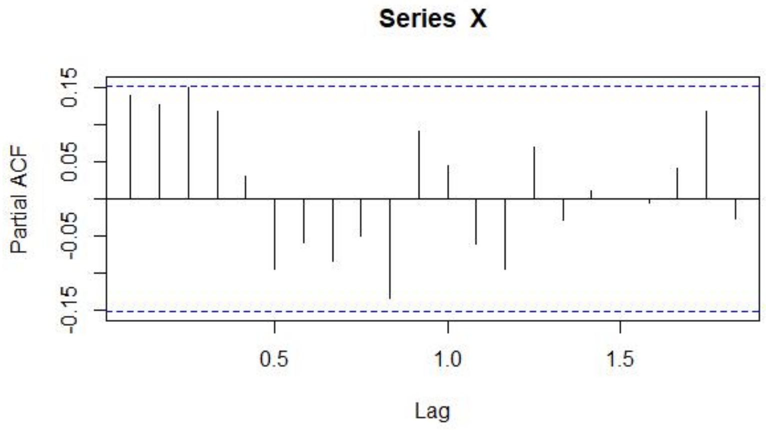
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -4.0963, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=0.

Fitted model by checking AIC is:

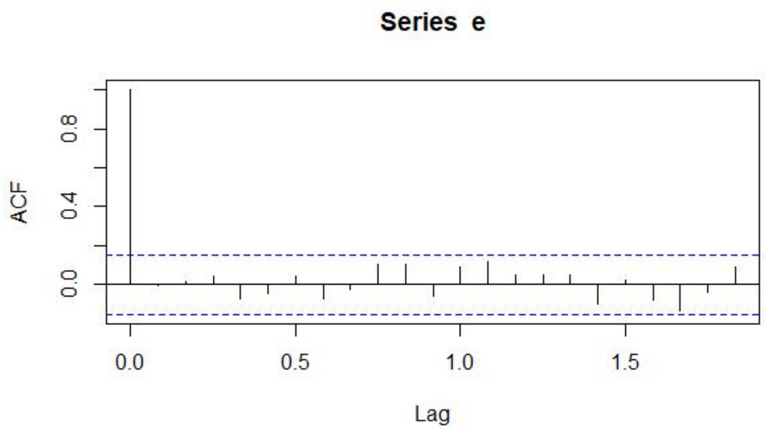
ARIMA(x = dat, order = c(1, 0, 2))

Coefficients:

	ar1	ma1	ma2	intercept
	0.9470	-0.9064	0.1316	17.1744
s.e.	0.0342	0.0824	0.0731	1.4059

sigma^2 estimated as 21.61: log likelihood = -496.86, aic = 1003.73

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0044439, df = 1, p-value = 0.9469

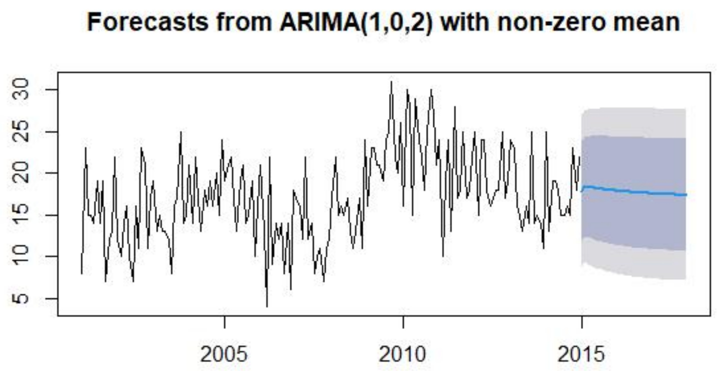
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	17.86223	18.41683	18.35101	18.28867	18.22964	18.17373	18.12079	18.07065
2016	17.85709	17.82093	17.78667	17.75423	17.72352	17.69442	17.66687	17.64078
2017	17.52965	17.51083	17.49301	17.47613	17.46014	17.44500	17.43067	17.41709

	Sep	Oct	Nov	Dec
2015	18.02317	17.97820	17.93562	17.89529
2016	17.61607	17.59267	17.57051	17.54953
2017	17.40423	17.39206	17.38053	17.36960

Note that, this plot is for deseasonalized data:



Now, the original predictions are, i.e. after adding the seasonal part is given below,

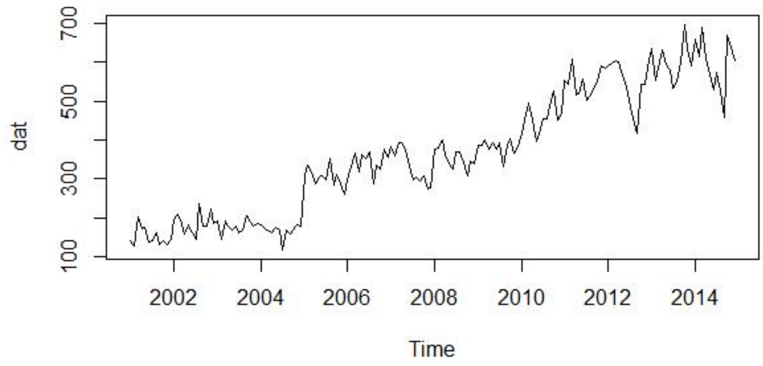
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	21.08240	20.73041	20.43162	17.76118	21.11757	18.92350	16.78875	17.42540
2016	21.79068	21.35721	20.84920	18.03938	21.30291	19.04698	16.87102	17.48020
2017	21.79788	21.36200	20.85239	18.04151	21.30433	19.04792	16.87165	17.48062

	Sep	Oct	Nov	Dec
2015	20.34768	20.76380	21.09570	22.67808
2016	20.38419	20.78813	21.11191	22.68887
2017	20.38447	20.78832	21.11203	22.68896

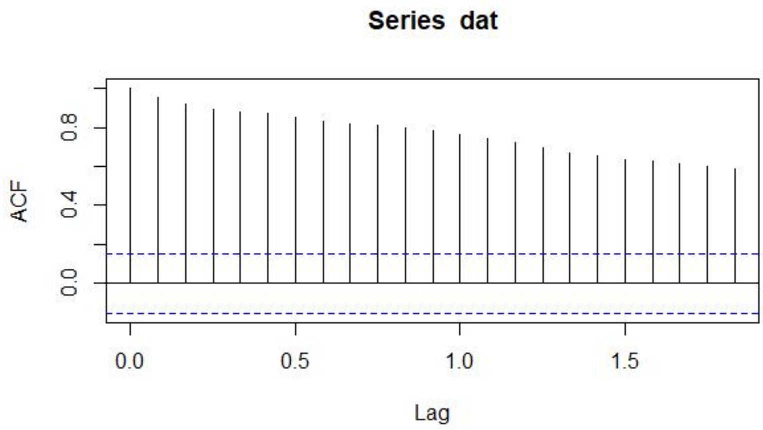
➢ **<u>Assam:</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



So, now we can say that the data is stationary,augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

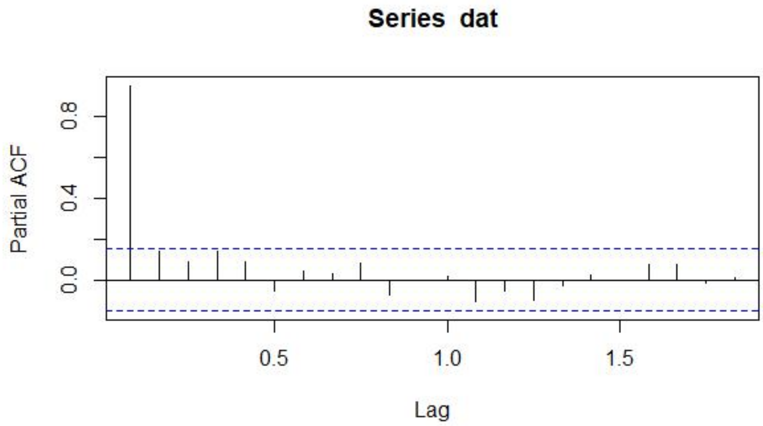
data: dat

Dickey-Fuller = -3.8694, Lag order = 5, p-value = 0.01728

alternative hypothesis: stationary

Therefore the data is stationary at d=0.

The pacf plot for the data:



Fitted model by checking AIC is:

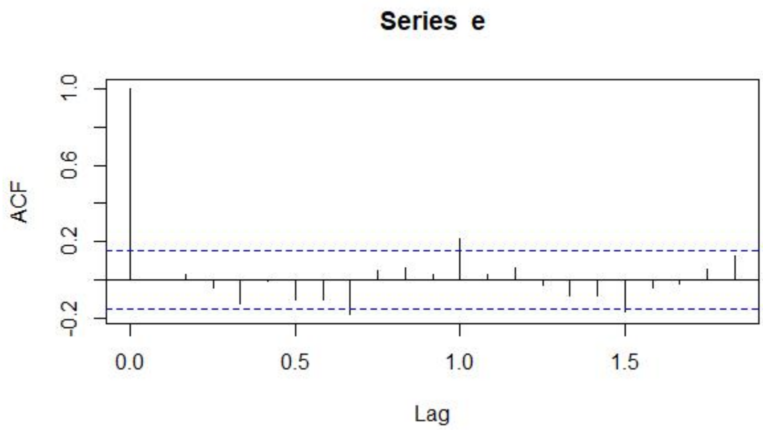
ARIMA(X,order = c(1,0,2)), so here we have taken p=1, d=0, q=2

Coefficients:

	ar1	ma1	ma2	intercept
	0.9947	-0.3394	-0.1648	379.1628
s.e.	0.0066	0.0782	0.0882	164.3946

sigma^2 estimated as 1656: log likelihood = -862.72, aic = 1735.44

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0020094, df = 1, p-value = 0.9642

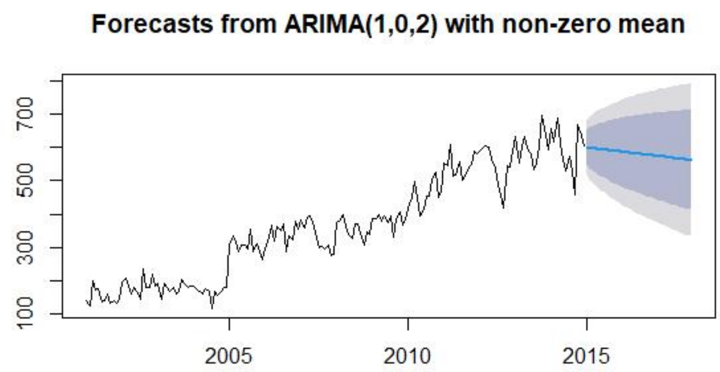
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	599.8201	598.4302	597.2718	596.1194	594.9732	593.8331	592.6989	591.5708
2016	586.0188	584.9259	583.8389	582.7575	581.6819	580.6120	579.5477	578.4890
2017	573.2790	572.2534	571.2333	570.2185	569.2092	568.2051	567.2064	566.2129

	Sep	Oct	Nov	Dec
2015	590.4486	589.3323	588.2220	587.1175
2016	577.4359	576.3884	575.3464	574.3100
2017	565.2247	564.2417	563.2639	562.2912

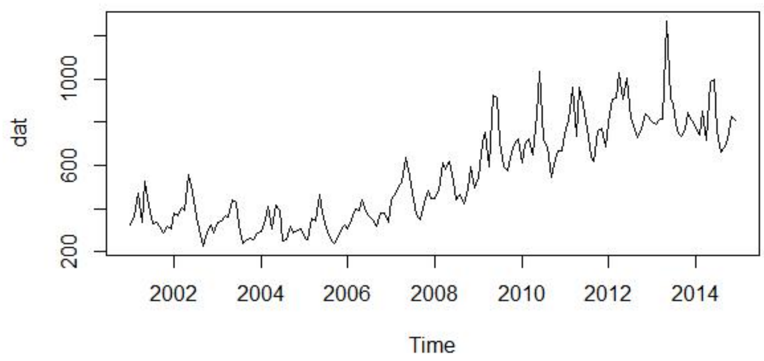
Plot for forecast is given below:



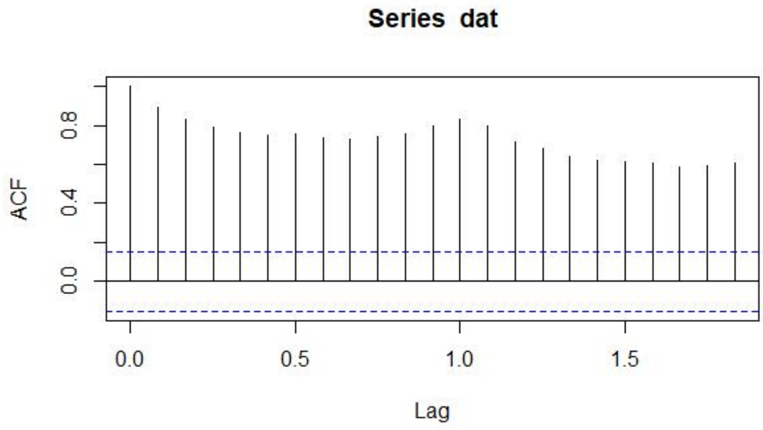
- **<u>Bihar:</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

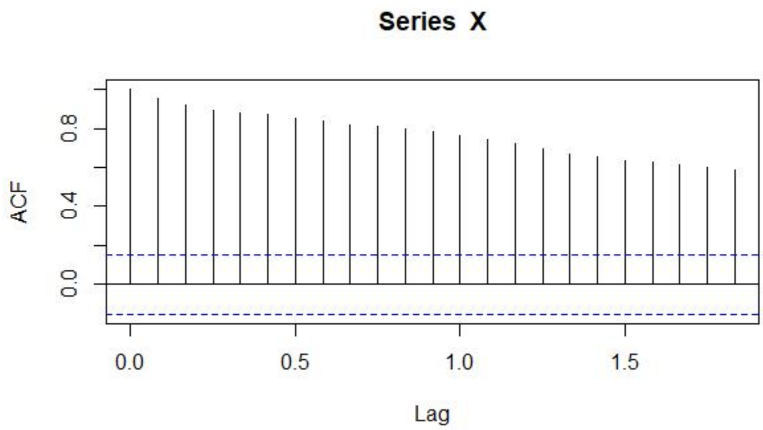


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.



So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

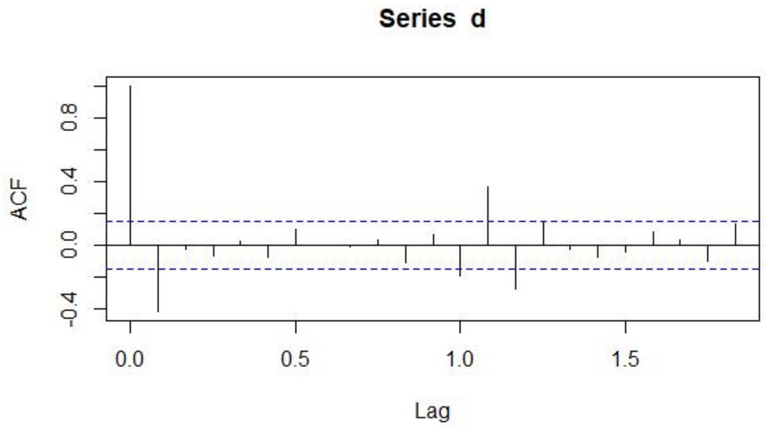
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -1.9146, Lag order = 5, p-value = 0.6119

alternative hypothesis: stationary

Now, the acf plot for the differenced data is given below



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

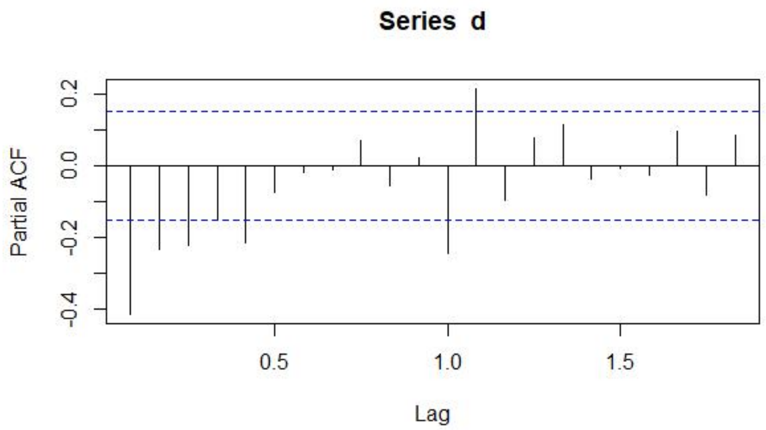
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -8.1598, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

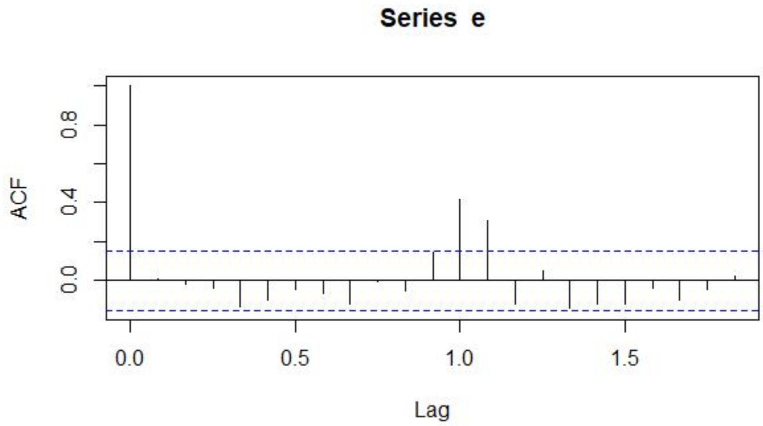
ARIMA(x = dat, order = c(2, 1,2))

Coefficients:

	ar1	ar2	ma1	ma2
	-0.5053	0.4942	0.1162	-0.8764
s.e.	0.0815	0.0816	0.0356	0.0353

sigma^2 estimated as 8471: log likelihood = -992.97, aic = 1995.95

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0088265, df = 1, p-value = 0.9251

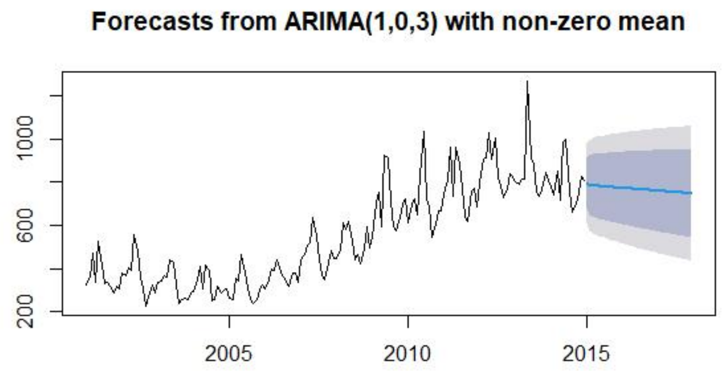
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	796.7450	784.5913	784.0753	782.8974	781.7261	780.5613	779.4028	778.2508
2016	772.5857	771.4714	770.3632	769.2612	768.1653	767.0754	765.9916	764.9138
2017	759.6134	758.5708	757.5340	756.5030	755.4776	754.4580	753.4439	752.4355

	Sep	Oct	Nov	Dec
2015	777.1052	775.9659	774.8329	773.7062
2016	763.8419	762.7760	761.7159	760.6617
2017	751.4326	750.4353	749.4435	748.4572

Note that, this plot is for deseasonalized data:



Now, the original predictions are, i.e. after adding the seasonal part is given below,

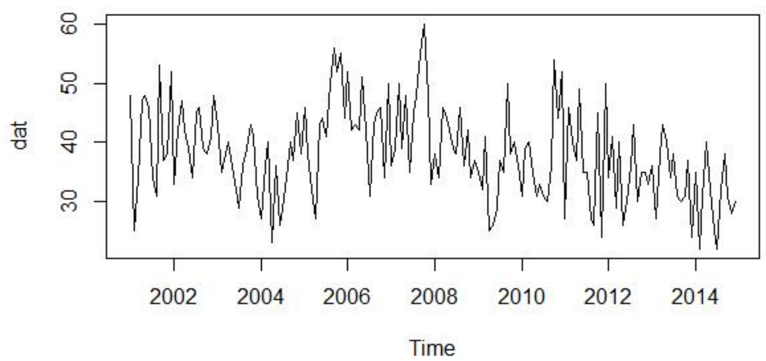
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	781.9964	800.0105	873.4020	805.5926	973.6621	911.3370	790.6771	720.6299
2016	775.3154	796.7557	871.7449	804.8214	973.2329	911.1728	790.5480	720.6140
2017	775.2820	796.7870	871.7126	804.8532	973.2009	911.2047	790.5161	720.6459

	Sep	Oct	Nov	Dec
2015	702.4190	727.8202	783.6809	743.8006
2016	702.3632	727.8405	783.6431	743.8298
2017	702.3313	727.8724	783.6112	743.8616

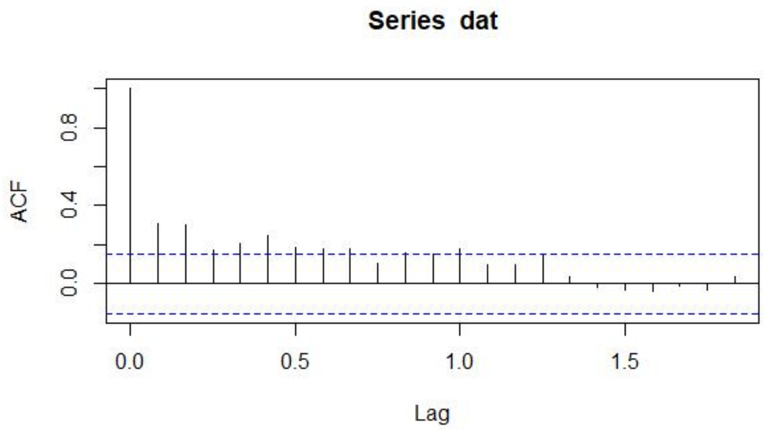
- **Chandigarh:**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is stationary.



So, now we can say that the data is stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

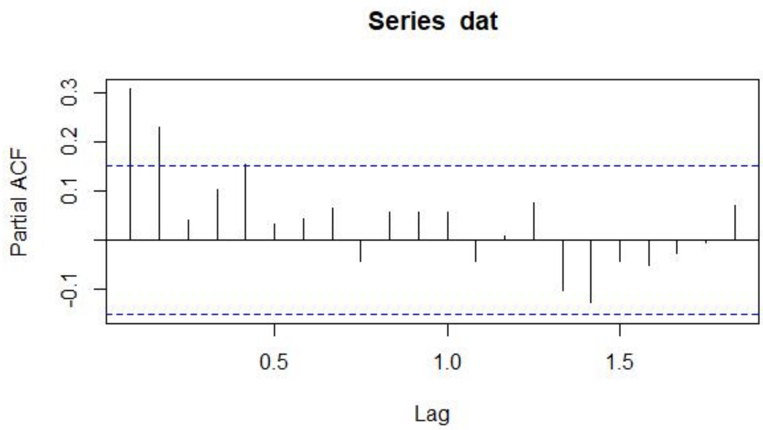
data: dat

Dickey-Fuller = -3.5134, Lag order = 5, p-value = 0.04333

alternative hypothesis: stationary

Therefore the data is stationary at d=0.

The pacf plot for the differenced data:



Fitted model by checking AIC is:

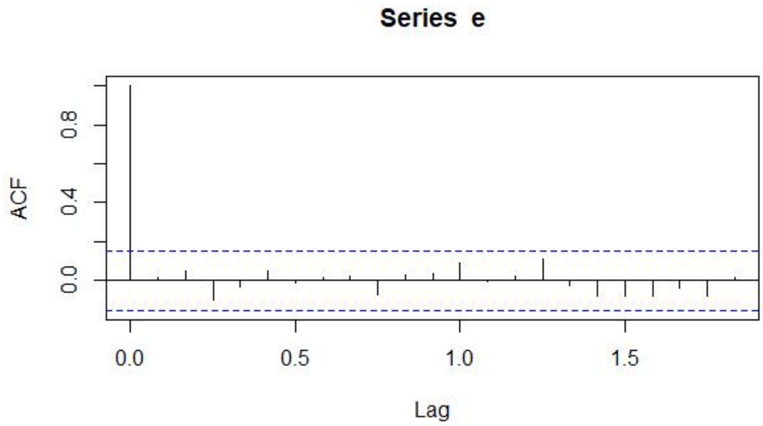
ARIMA(X,order = c(1,0,1)), so here we have taken p=1, d=0, q=1

Coefficients:

	ar1	ma1	intercept
	0.9287	-0.7583	37.8386
s.e.	0.0541	0.0944	1.7797

sigma^2 estimated as 50.77: log likelihood = -568.48, aic = 1144.96

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.065035, df = 1, p-value = 0.7987

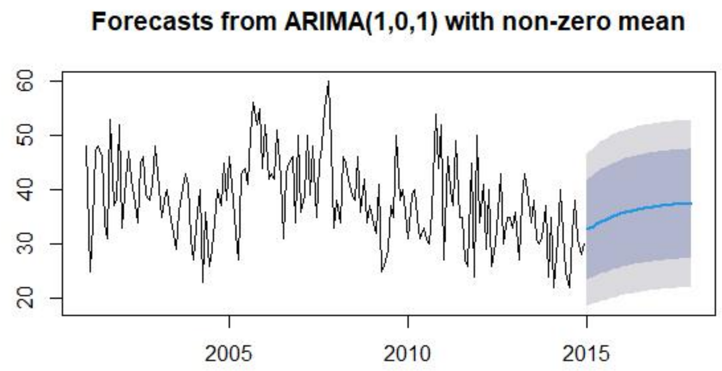
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	32.60580	32.97916	33.32588	33.64787	33.94687	34.22455	34.48241	34.72188
2016	35.68603	35.83962	35.98224	36.11470	36.23770	36.35193	36.45800	36.55651
2017	36.95313	37.01631	37.07498	37.12947	37.18007	37.22706	37.27069	37.31121

	Sep	Oct	Nov	Dec
2015	34.94426	35.15077	35.34255	35.52064
2016	36.64799	36.73294	36.81183	36.88509
2017	37.34885	37.38379	37.41624	37.44638

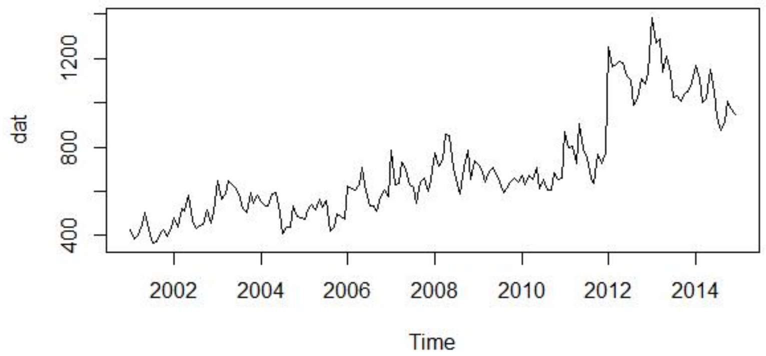
Plot for forecast is given below:



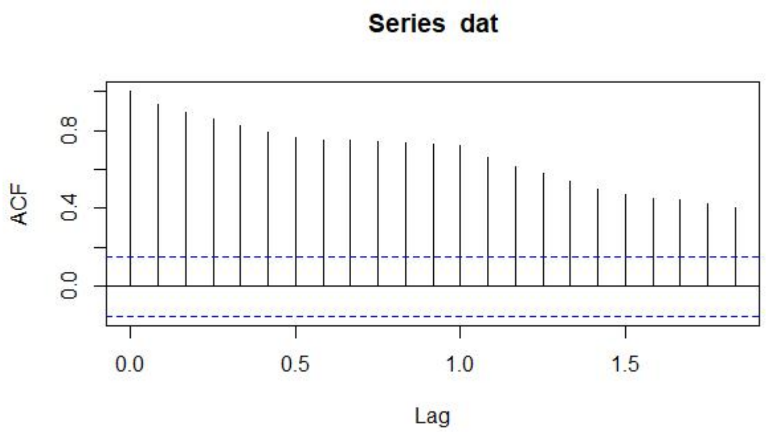
➢ **<u>Chhattisgarh:</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

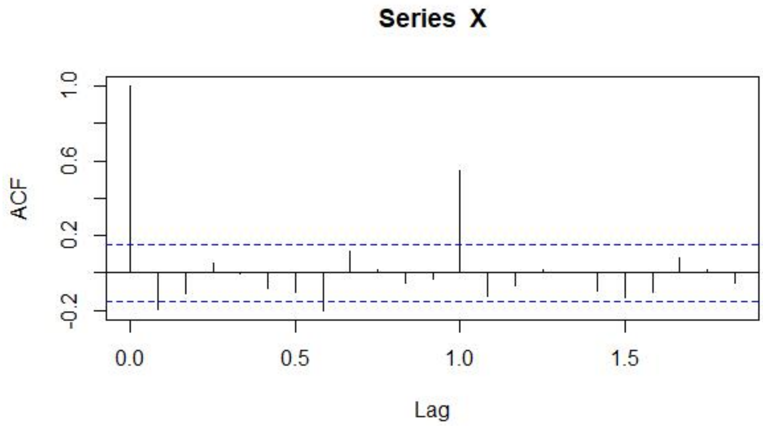
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -3.1287, Lag order = 5, p-value = 0.1052

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

Augmented Dickey-Fuller Test

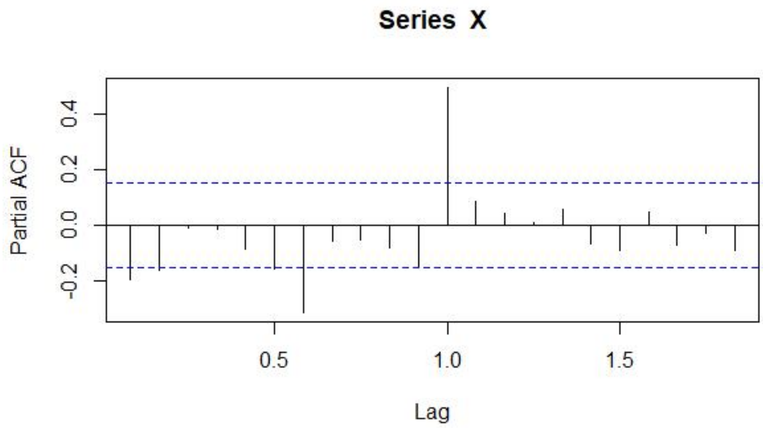
data: X

Dickey-Fuller = -6.9056, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

Therefore the data is stationary at d=1.

The pacf plot for the differenced data:



Fitted model by checking AIC is:

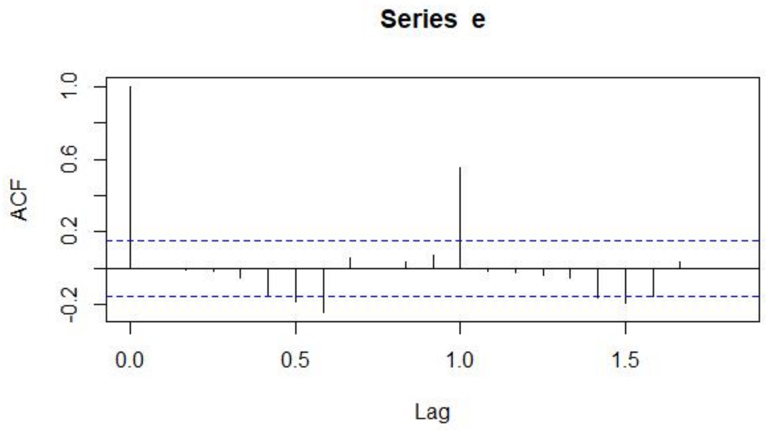
ARIMA(X,order = c(2,1,1)), so here we have taken p=2, d=1, q=1

Coefficients:

	ar1	ar2	ma1
	-0.1816	-0.1463	-0.0414
s.e.	0.5780	0.1371	0.5858

sigma^2 estimated as 6090: log likelihood = -964.66, aic = 1937.32

For diagnostic checking for model, we have used the residual’s acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0030536, df = 1, p-value = 0.9559

There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

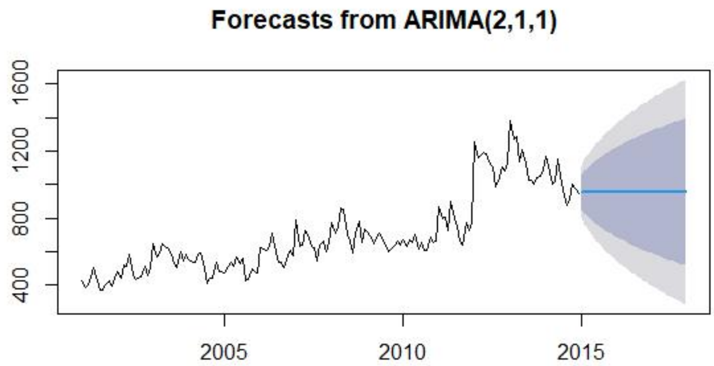
Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	954.8238	957.1022	955.1046	955.1339	955.4209	955.3645	955.3327	955.3468
2016	955.3468	955.3468	955.3468	955.3468	955.3468	955.3468	955.3468	955.3468
2017	955.3468	955.3468	955.3468	955.3468	955.3468	955.3468	955.3468	955.3468

	Sep	Oct	Nov	Dec
2015	955.3489	955.3464	955.3466	955.3469
2016	955.3468	955.3468	955.3468	955.3468

2017 955.3468 955.3468 955.3468 955.3468

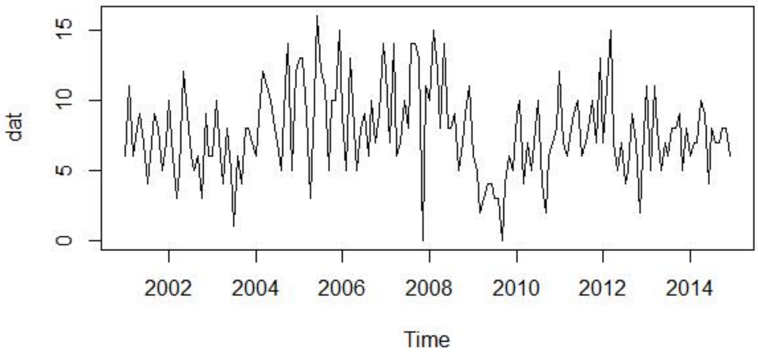
Plot for forecast is given below:



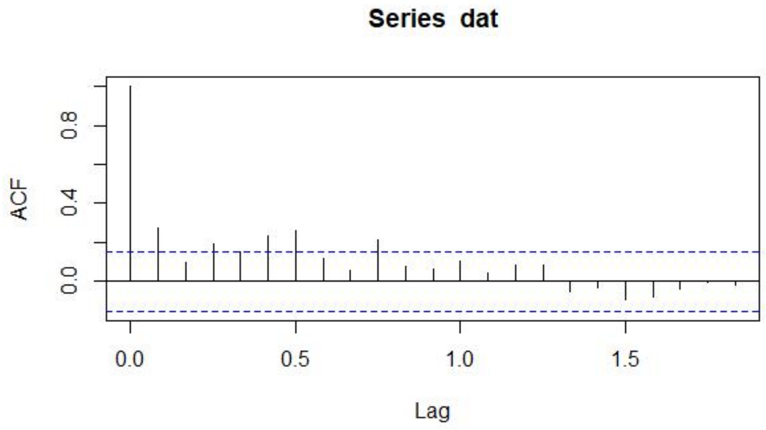
➢ **<u>Dadar and Nagar Haveli :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

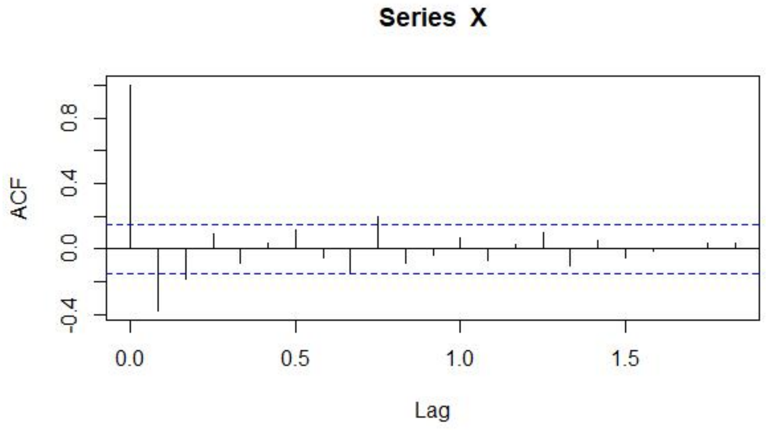
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -2.9242, Lag order = 5, p-value = 0.1905

alternative hypothesis: stationary

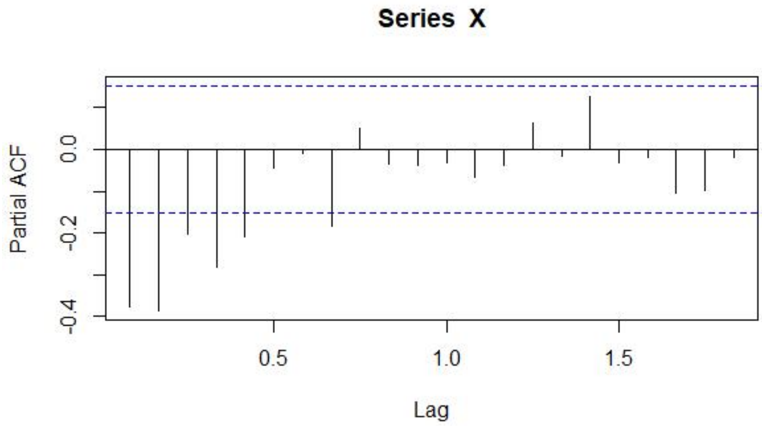
Now, we will take the diff to make it stationary.

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -8.365, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary



Therefore the data is stationary at d=1.

The pacf plot for the differenced data:

Fitted model by checking AIC is:

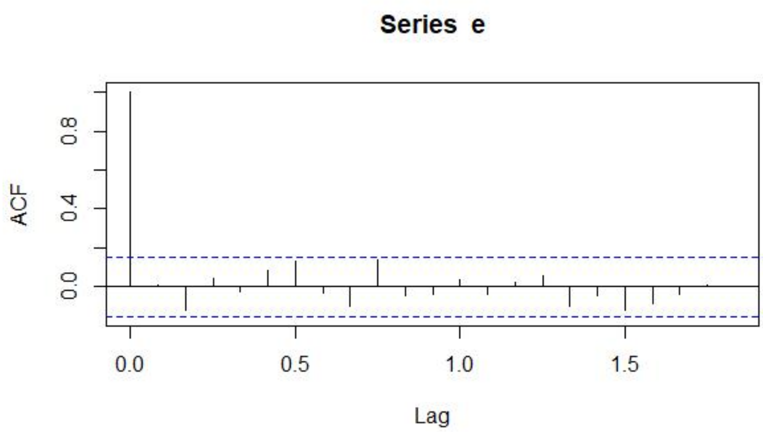
ARIMA(X,order = c(1,1,1)), so here we have taken p=1, d=1, q=1

Coefficients:

	ar1	ma1
	0.1093	-0.8707
s.e.	0.0950	0.0523

sigma^2 estimated as 9.131: log likelihood = -422.25, aic = 850.51

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.016817, df = 1, p-value = 0.8968

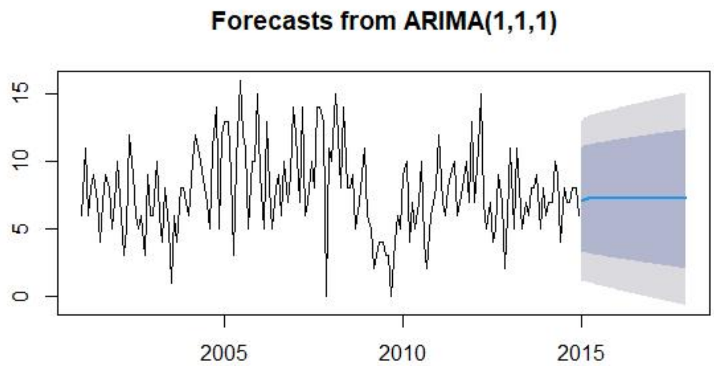
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	7.074537	7.191992	7.204831	7.206234	7.206388	7.206405	7.206406	7.206407
2016	7.206407	7.206407	7.206407	7.206407	7.206407	7.206407	7.206407	7.206407
2017	7.206407	7.206407	7.206407	7.206407	7.206407	7.206407	7.206407	7.206407

	Sep	Oct	Nov	Dec
2015	7.206407	7.206407	7.206407	7.206407
2016	7.206407	7.206407	7.206407	7.206407
2017	7.206407	7.206407	7.206407	7.206407

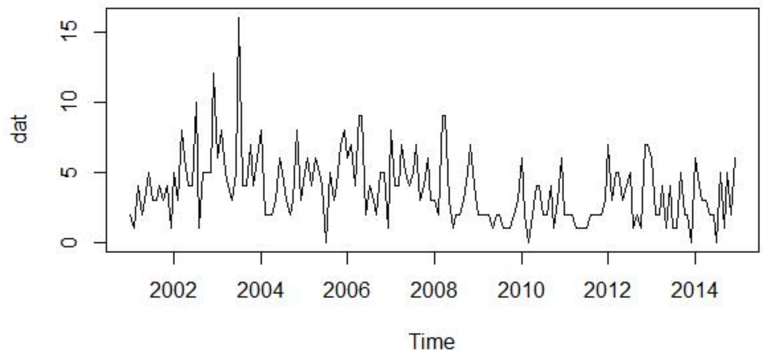
Plot for forecast is given below:



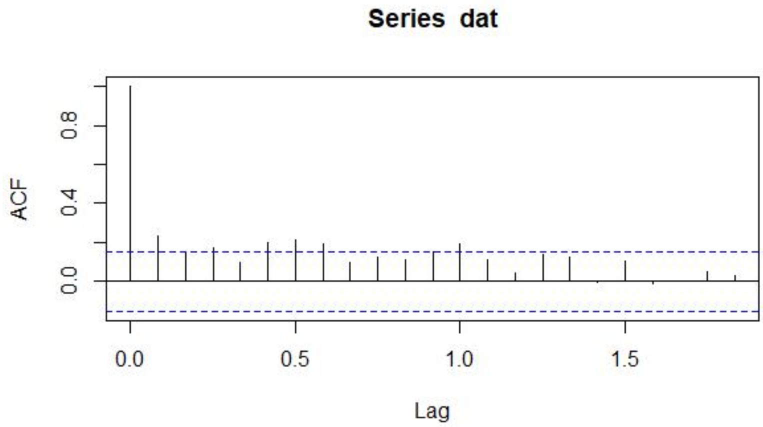
- **<u>Daman and Diu :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is stationary and may be there is seasonality present.



So, now we can say that the data is stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

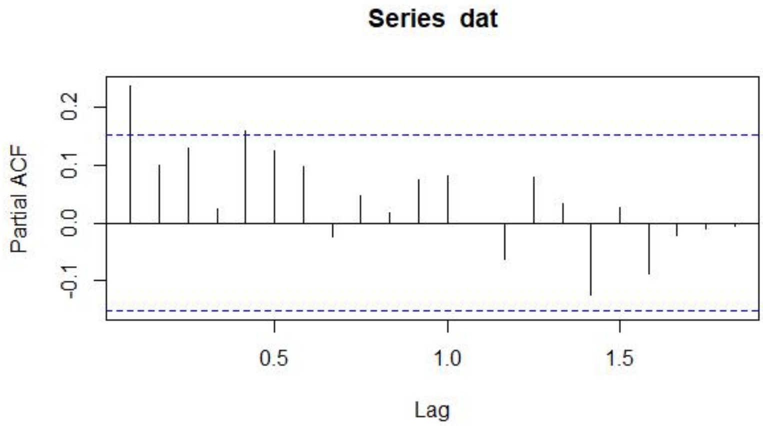
data: dat

Dickey-Fuller = -4.1398, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

Therefore the data is stationary at d=0.

The pacf plot for the data:



Fitted model by checking AIC is:

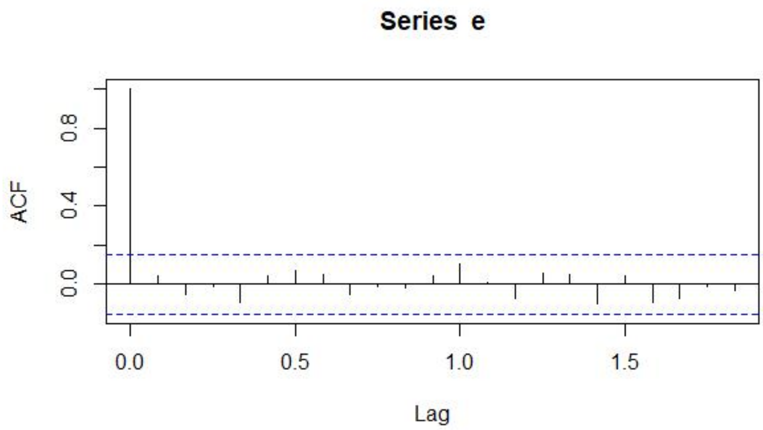
ARIMA(X,order = c(1,0,1)), so here we have taken p=1, d=0, q=1

Coefficients:

	ar1	ma1	intercept
	0.9468	-0.8372	3.7381
s.e.	0.0430	0.0748	0.5124

sigma^2 estimated as 5.326: log likelihood = -379.05, aic = 766.09

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.30689, df = 1, p-value = 0.5796

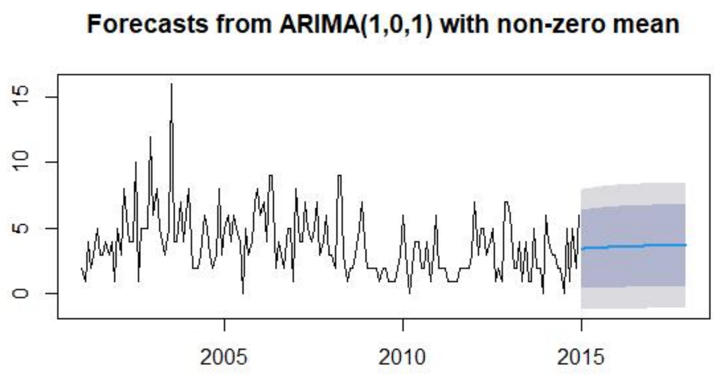
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3.414787	3.431984	3.448267	3.463683	3.478280	3.492101	3.505186	3.517576
2016	3.570313	3.579239	3.587690	3.595692	3.603268	3.610441	3.617233	3.623664
2017	3.651036	3.655668	3.660055	3.664208	3.668140	3.671863	3.675388	3.678726

	Sep	Oct	Nov	Dec
2015	3.529306	3.540413	3.550929	3.560886
2016	3.629752	3.635517	3.640975	3.646143
2017	3.681886	3.684878	3.687711	3.690393

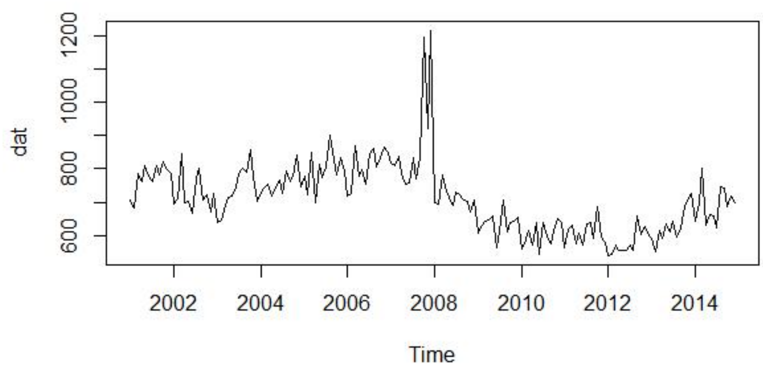
Plot for forecast is given below:



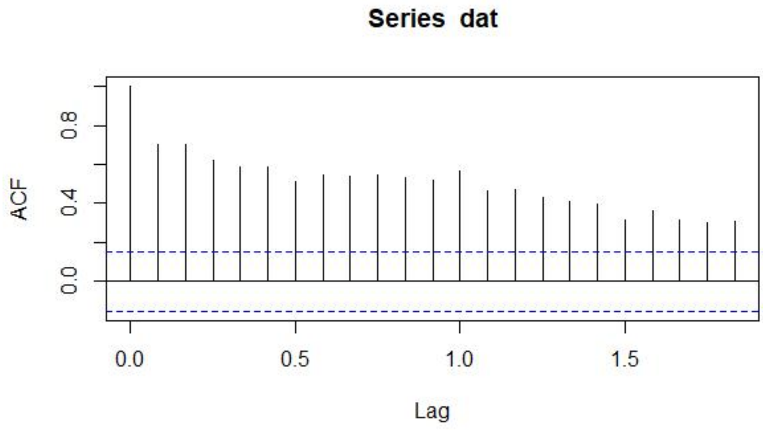
➢ **Delhi (UT) :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

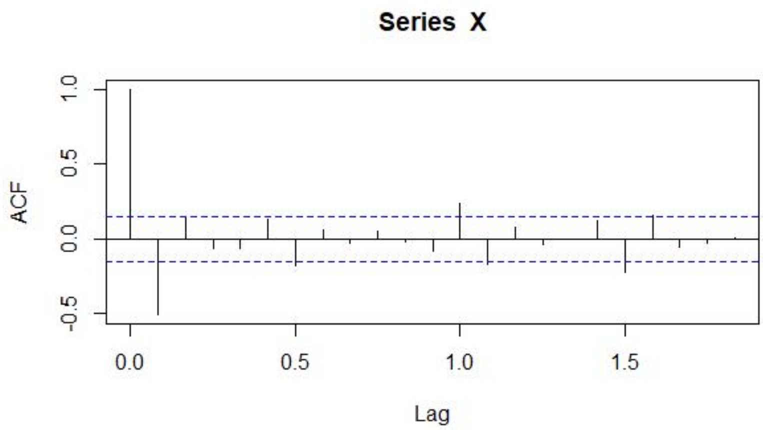
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -2.6283, Lag order = 5, p-value = 0.314

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

Augmented Dickey-Fuller Test

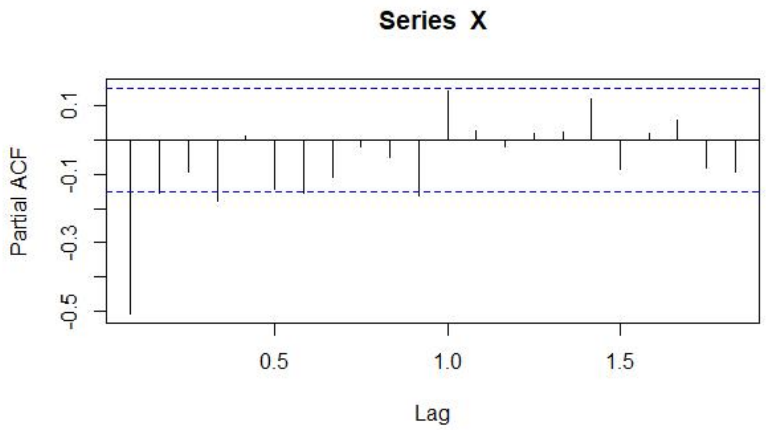
data: X

Dickey-Fuller = -7.2375, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

Therefore the data is stationary at d=1.

The pacf plot for the differenced data:



Fitted model by checking AIC is:

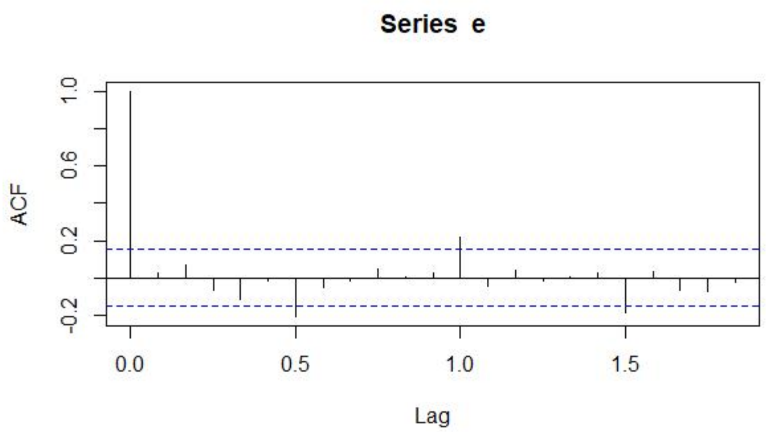
ARIMA(X,order = c(1,1,1)), so here we have taken p=1, d=1, q=1

Coefficients:

	ar1	ma1
	0.0533	-0.7000
s.e.	0.1561	0.1324

sigma^2 estimated as 4666: log likelihood = -942.67, aic = 1891.34

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.01506, df = 1, p-value = 0.9023

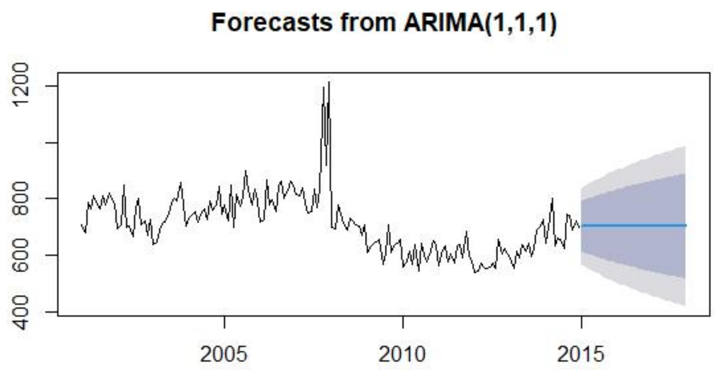
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	702.6642	702.8062	702.8137	702.8141	702.8142	702.8142	702.8142	702.8142
2016	702.8142	702.8142	702.8142	702.8142	702.8142	702.8142	702.8142	702.8142
2017	702.8142	702.8142	702.8142	702.8142	702.8142	702.8142	702.8142	702.8142

	Sep	Oct	Nov	Dec
2015	702.8142	702.8142	702.8142	702.8142
2016	702.8142	702.8142	702.8142	702.8142
2017	702.8142	702.8142	702.8142	702.8142

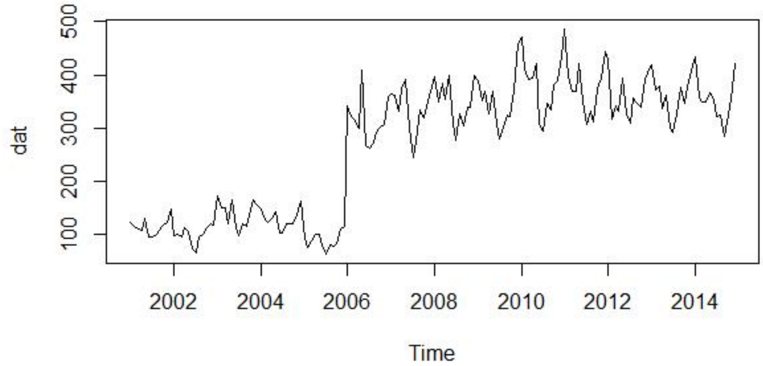
Plot for forecast is given below:



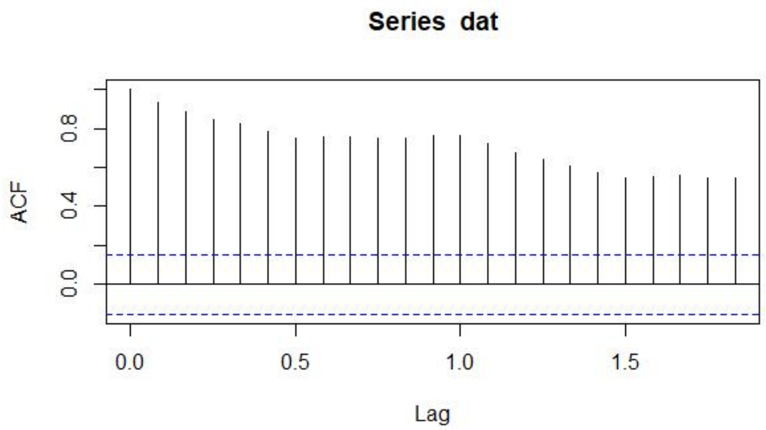
- **<u>Goa:</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

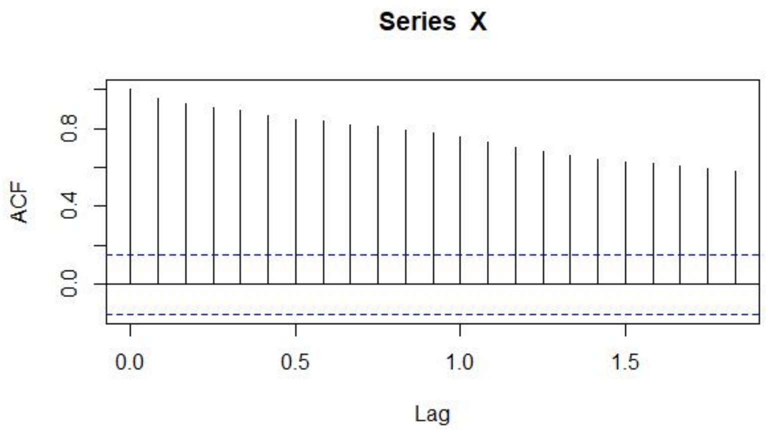


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

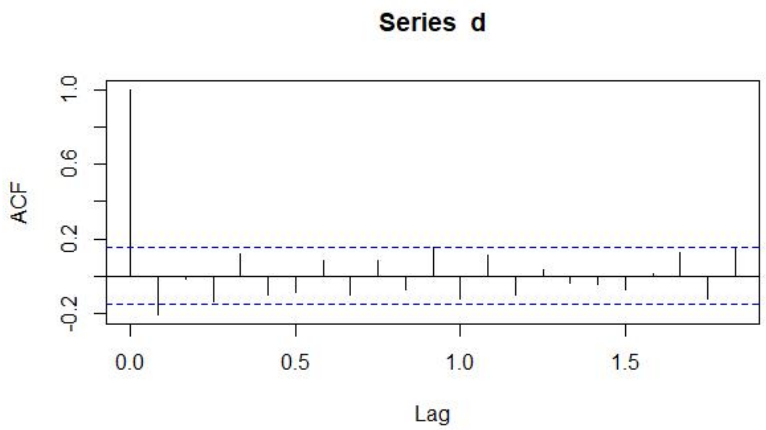


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data’s acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -1.9702, Lag order = 5, p-value = 0.5887

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

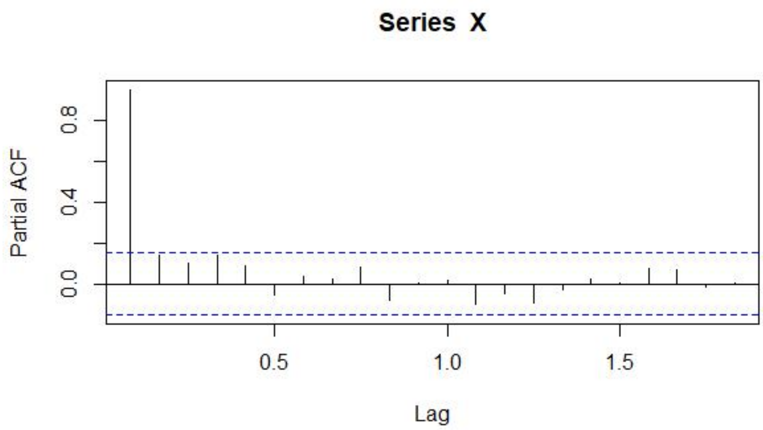
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -6.8122, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

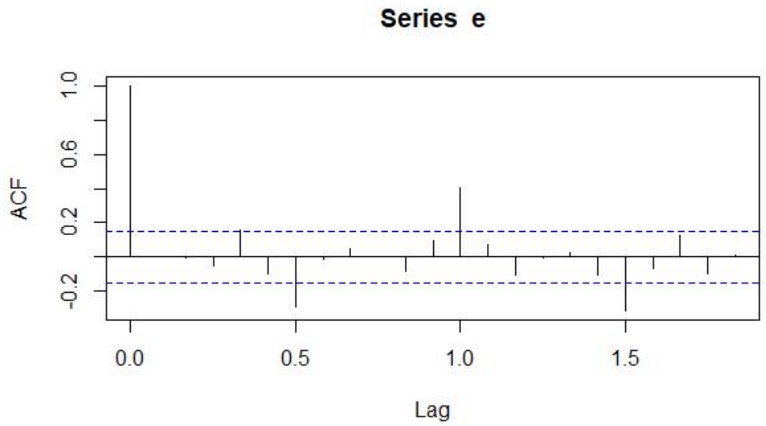
ARIMA(x = dat, order = c(1, 1, 1)), so here we have taken p=1, d=1, q=1

Coefficients:

	ar1	ma1
	0.6473	-0.8659
s.e.	0.0994	0.0602

sigma^2 estimated as 1633: log likelihood = -854.83, aic = 1715.66

For diagnostic checking for model, we have used the residual’s acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.00041243, df = 1, p-value = 0.9838

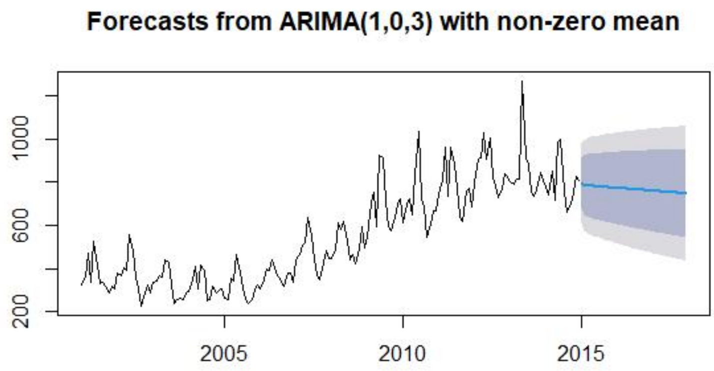
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	404.4038	393.0132	385.6398	380.8667	377.7770	375.7769	374.4822	373.6441
2016	372.2806	372.2190	372.1791	372.1532	372.1365	372.1257	372.1187	372.1141
2017	372.1067	372.1064	372.1062	372.1060	372.1060	372.1059	372.1059	372.1058

	Sep	Oct	Nov	Dec
2015	373.1016	372.7504	372.5231	372.3759
2016	372.1112	372.1093	372.1081	372.1073
2017	372.1058	372.1058	372.1058	372.1058

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

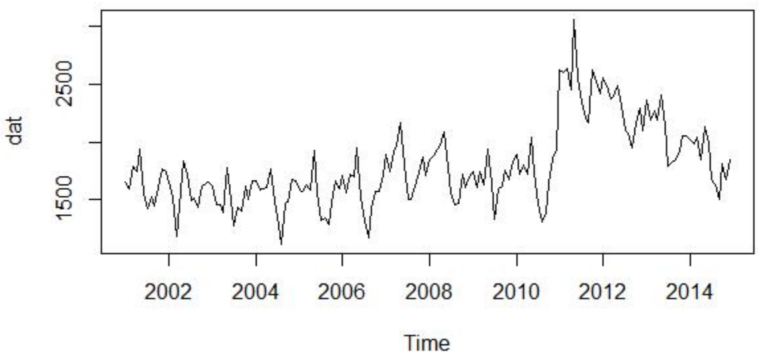
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	458.8934	404.2913	395.0204	382.0550	412.1352	342.8979	319.2667	344.5023
2016	426.7702	383.4970	381.5597	373.3415	406.4947	339.2467	316.9031	342.9723
2017	426.5963	383.3844	381.4868	373.2943	406.4641	339.2269	316.8903	342.9640

	Sep	Oct	Nov	Dec
2015	349.7995	359.9131	379.1569	408.4232
2016	348.8091	359.2719	378.7419	408.1545
2017	348.8037	359.2685	378.7396	408.1531

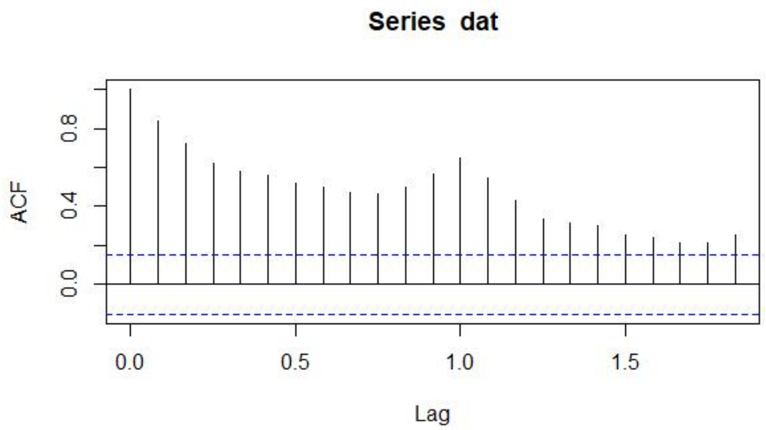
- **<u>Gujarat:</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

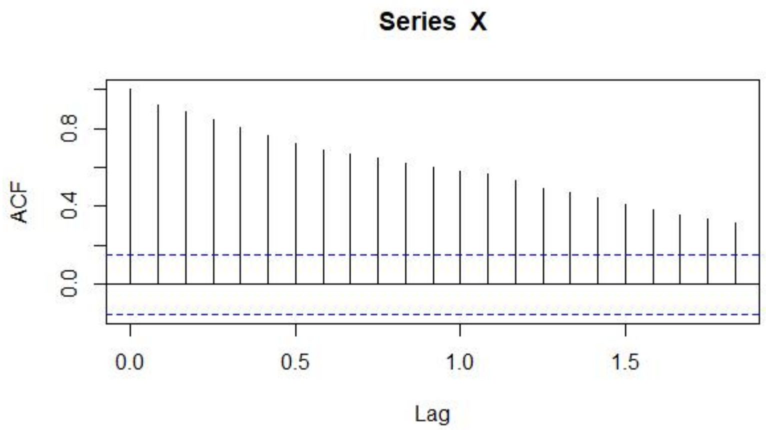


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

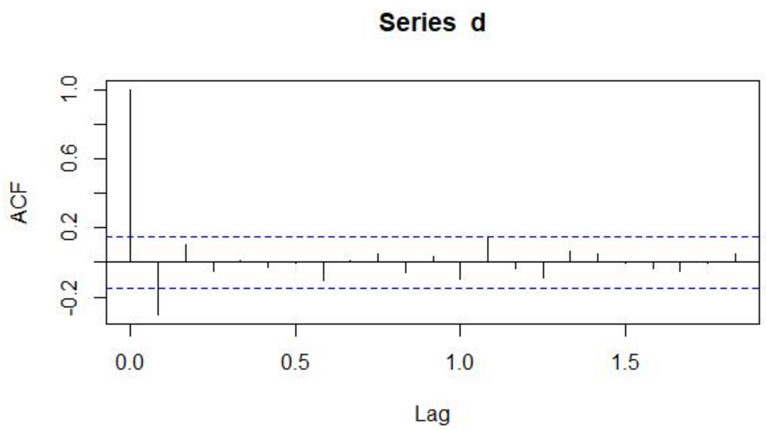


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -2.4429, Lag order = 5, p-value = 0.3914

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

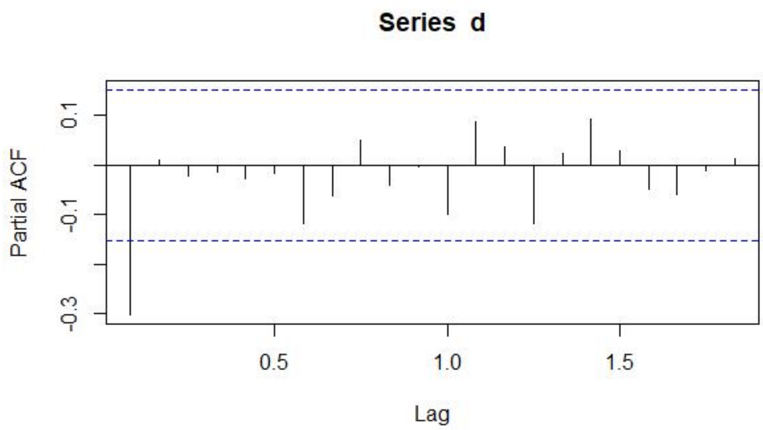
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -5.5111, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

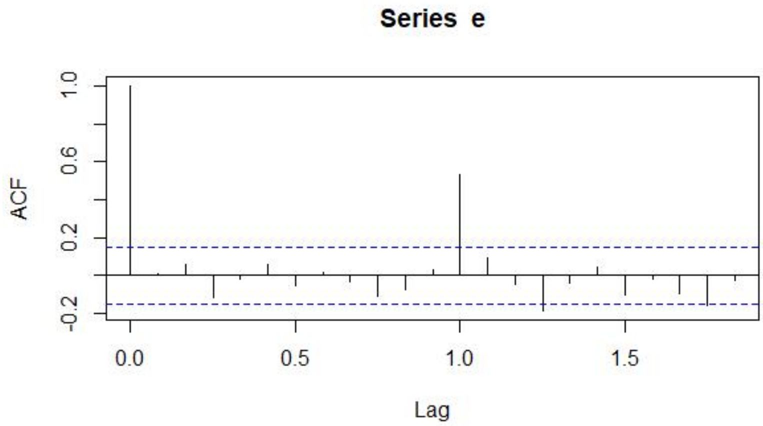
ARIMA(x = dat, order = c(1, 1, 1)), so here we have taken p=1, d=1, q=1

Coefficients:

	ar1	ma1
	0.6482	-0.9006
s.e.	0.1072	0.0658

sigma^2 estimated as 34857: log likelihood = -1110.52, aic = 2227.04

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0064501, df = 1, p-value = 0.936

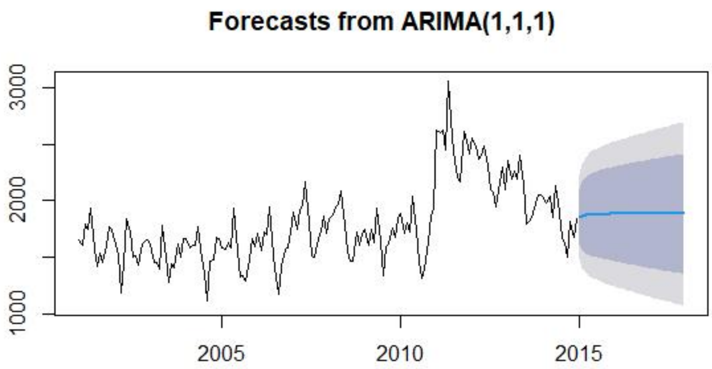
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	1856.958	1865.357	1870.801	1874.330	1876.617	1878.099	1879.060	1879.683
2016	1880.700	1880.746	1880.776	1880.795	1880.808	1880.816	1880.821	1880.825
2017	1880.830	1880.830	1880.831	1880.831	1880.831	1880.831	1880.831	1880.831

	Sep	Oct	Nov	Dec
2015	1880.087	1880.349	1880.518	1880.628
2016	1880.827	1880.828	1880.829	1880.830
2017	1880.831	1880.831	1880.831	1880.831

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

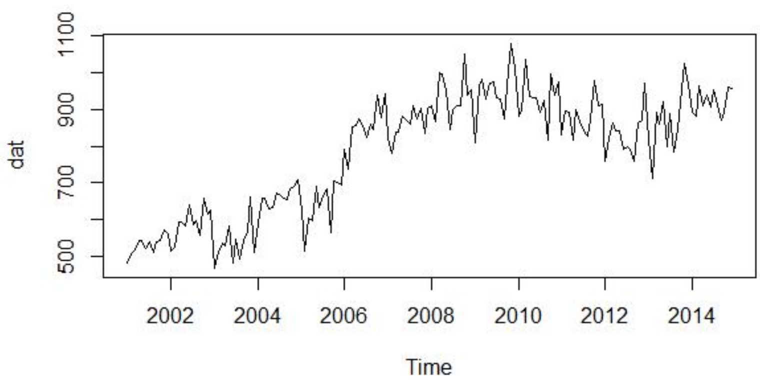
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	1996.055	1905.630	1942.183	1913.145	2199.188	1905.485	1655.437	1637.752
2016	2019.797	1921.019	1952.158	1919.610	2203.379	1908.202	1657.197	1638.893
2017	2019.927	1921.104	1952.213	1919.646	2203.402	1908.217	1657.207	1638.899

	Sep	Oct	Nov	Dec
2015	1668.918	1851.674	1911.593	1915.428
2016	1669.658	1852.153	1911.904	1915.629
2017	1669.662	1852.156	1911.906	1915.630

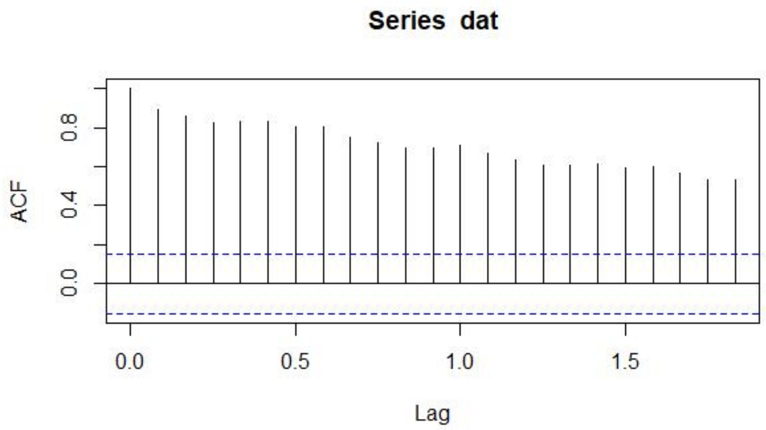
- **<u>Haryana :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

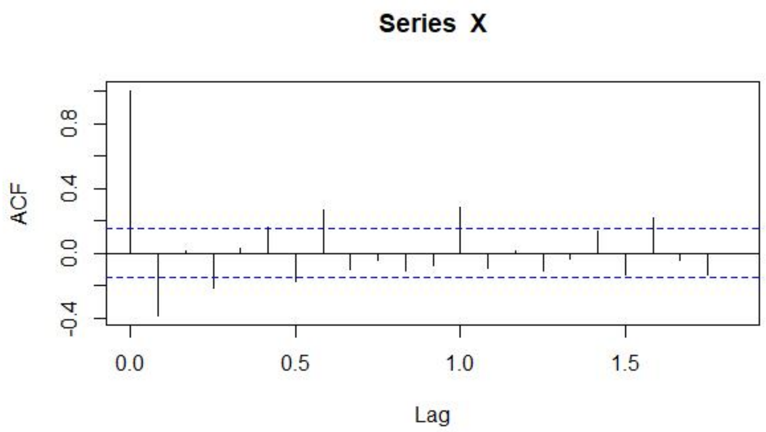
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -1.5395, Lag order = 5, p-value = 0.7685

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

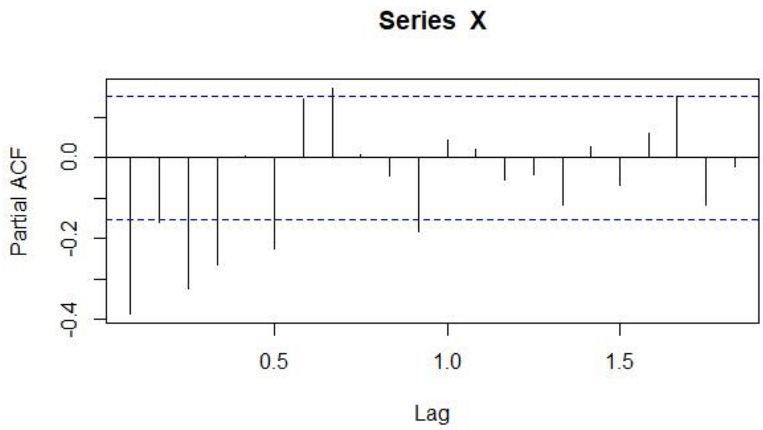
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -8.8156, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

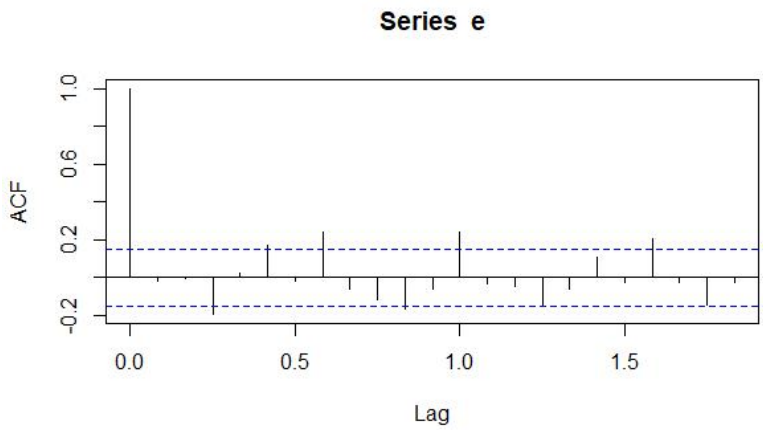
ARIMA(X,order = c(1,1,2)), so here we have taken p=1, d=1, q=2

Coefficients:

	ar1	ma1	ma2
	0.1051	-0.6754	-0.0292
s.e.	0.3390	0.3322	0.2168

sigma^2 estimated as 3597: log likelihood = -920.91, aic = 1849.83

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.062638, df = 1, p-value = 0.8024

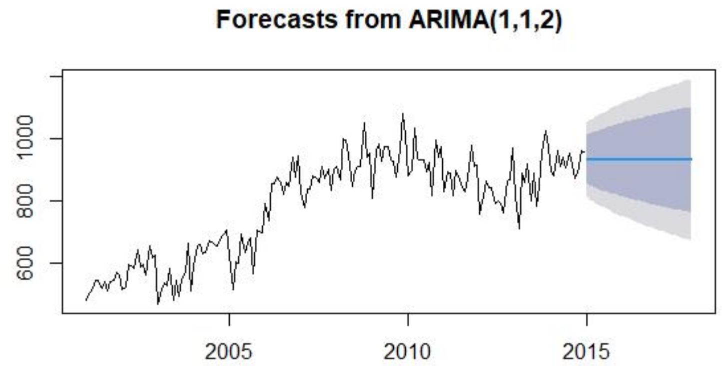
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	935.7678	932.8687	932.5640	932.5320	932.5286	932.5282	932.5282	932.5282
2016	932.5282	932.5282	932.5282	932.5282	932.5282	932.5282	932.5282	932.5282
2017	932.5282	932.5282	932.5282	932.5282	932.5282	932.5282	932.5282	932.5282

	Sep	Oct	Nov	Dec
2015	932.5282	932.5282	932.5282	932.5282
2016	932.5282	932.5282	932.5282	932.5282
2017	932.5282	932.5282	932.5282	932.5282

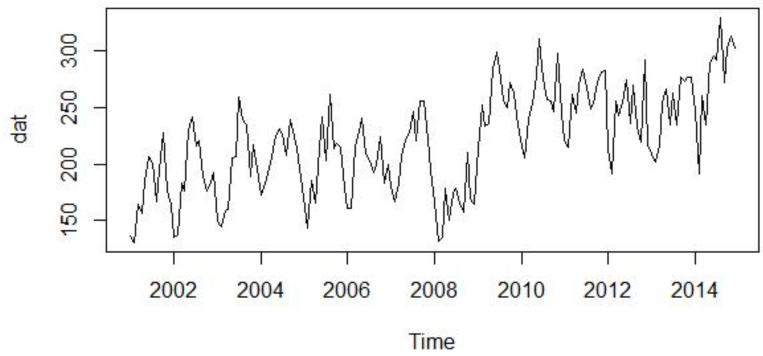
Plot for forecast is given below:



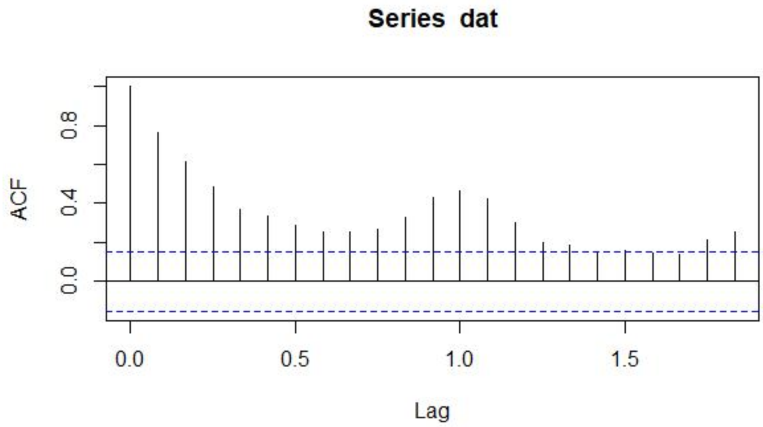
- **<u>Himachal Pradesh :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

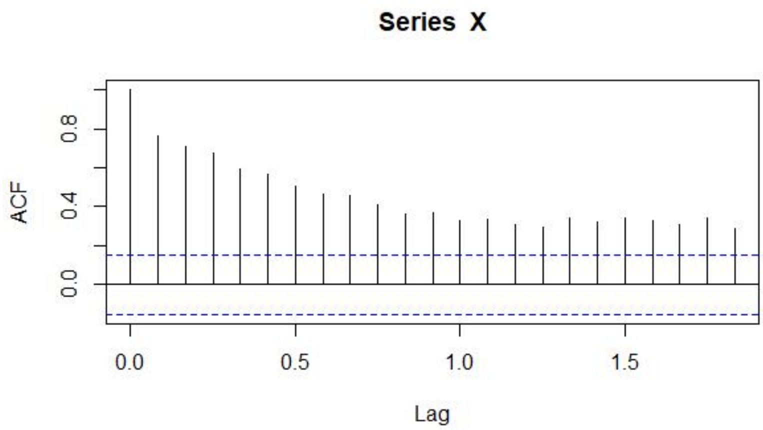


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

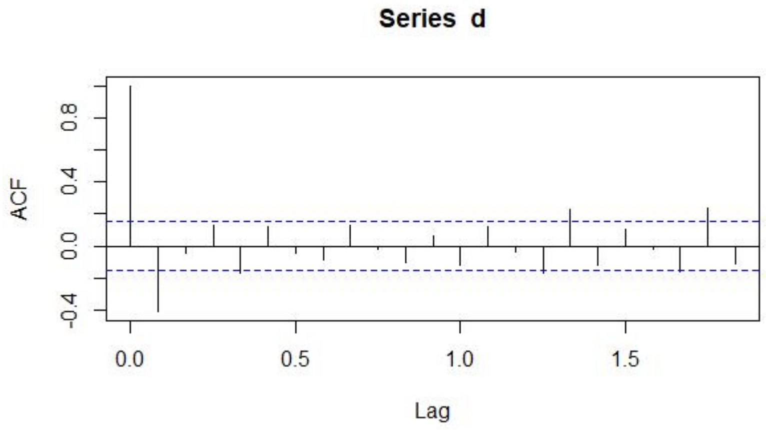


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -3.3781, Lag order = 5, p-value = 0.06049

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

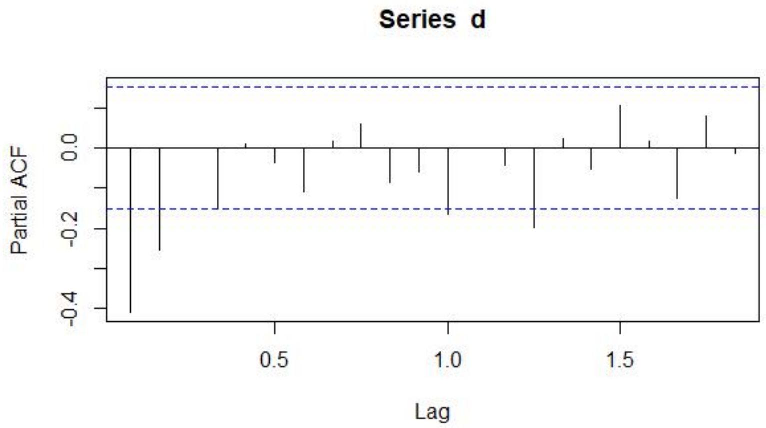
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -6.2948, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

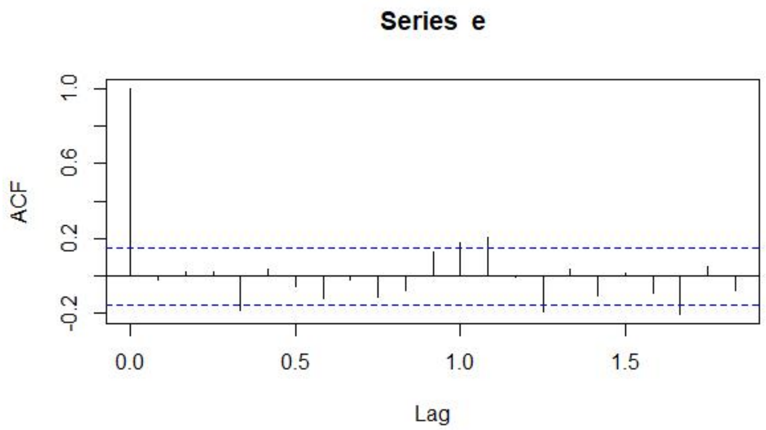
ARIMA(x = dat, order = c(2, 1, 2)), so here we have taken p=2, d=1, q=2

Coefficients:

	ar1	ar2	ma1	ma2
	0.4078	0.1679	-0.7475	-0.1792
s.e.	0.5247	0.3263	0.5252	0.4891

sigma^2 estimated as 731.7: log likelihood = -788.07, aic = 1586.14

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.059918, df = 1, p-value = 0.8066

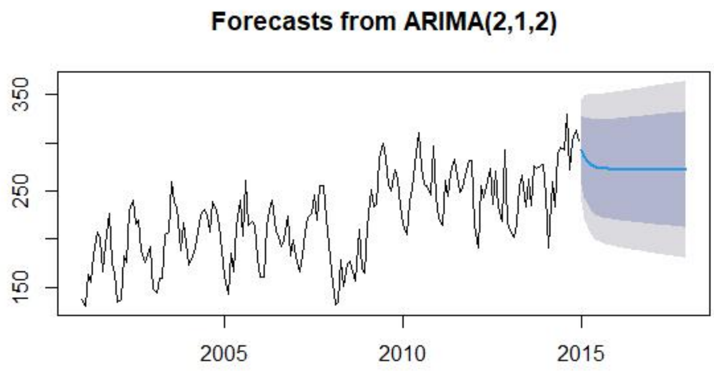
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	291.9065	285.2434	280.8311	277.9129	275.9819	274.7044	273.8591	273.2999
2016	272.3451	272.2982	272.2671	272.2466	272.2330	272.2240	272.2181	272.2141
2017	272.2074	272.2071	272.2069	272.2067	272.2066	272.2066	272.2065	272.2065

	Sep	Oct	Nov	Dec
2015	272.9299	272.6851	272.5231	272.4160
2016	272.2115	272.2098	272.2087	272.2079
2017	272.2065	272.2065	272.2065	272.2065

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

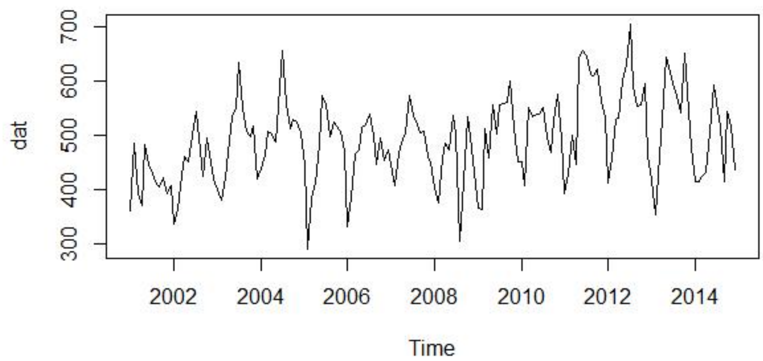
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	260.0563	243.1977	268.0964	270.3063	293.5420	299.0786	292.3872	282.3632
2016	240.4949	230.2525	259.5324	264.6400	289.7931	296.5982	290.7461	281.2774
2017	240.3573	230.1614	259.4721	264.6002	289.7667	296.5808	290.7346	281.2698

	Sep	Oct	Nov	Dec
2015	282.1855	286.0753	283.5576	263.4472
2016	281.4672	285.6000	283.2431	263.2392
2017	281.4621	285.5967	283.2409	263.2377

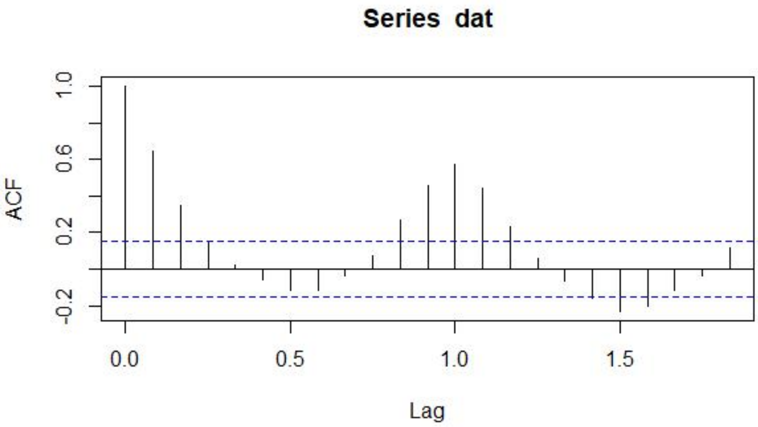
- **<u>Jammu and Kashmir :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

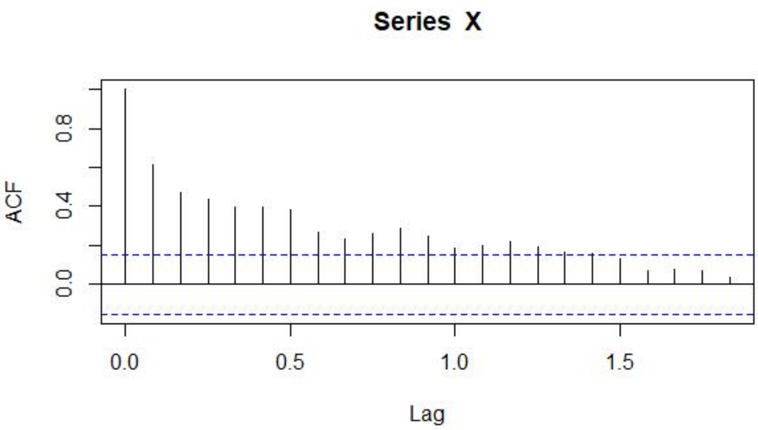


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

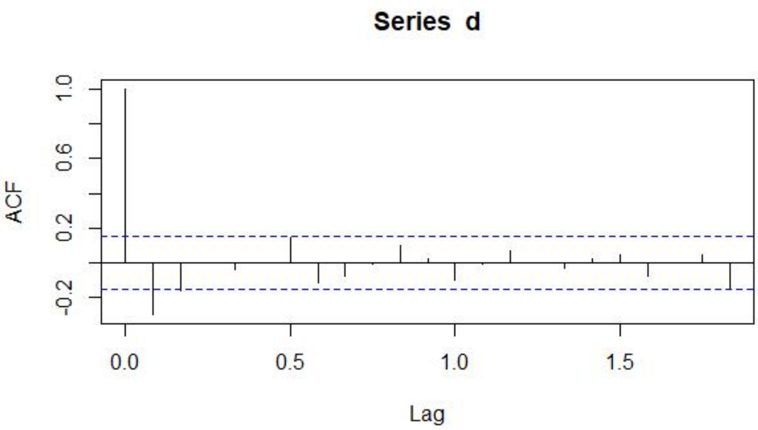


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -2.8896, Lag order = 5, p-value = 0.205

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

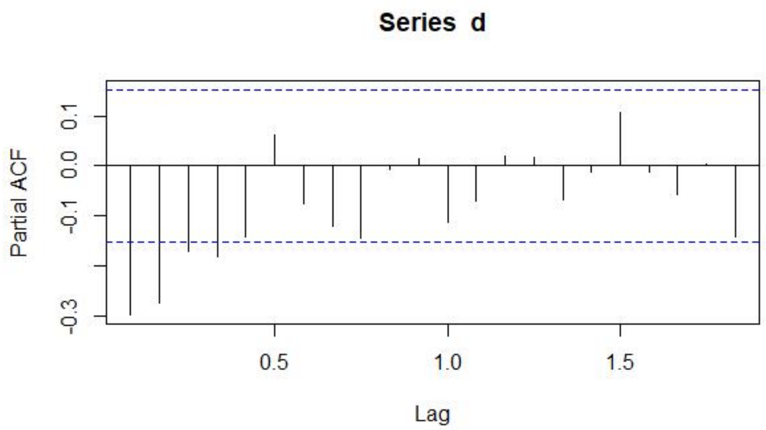
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -7.0389, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

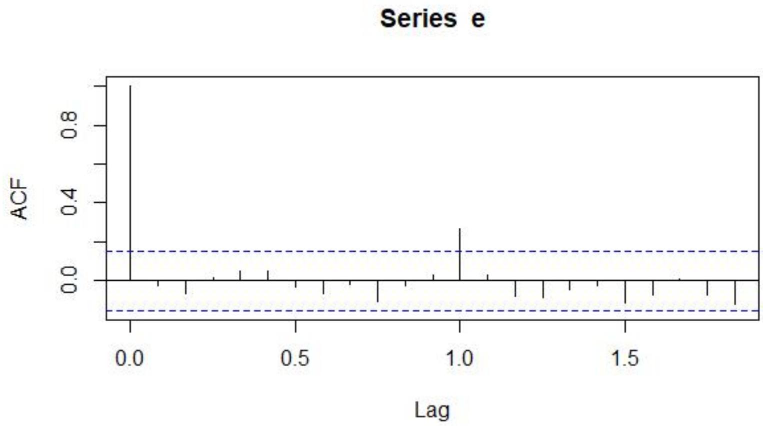
Fitted model by checking AIC is:

ARIMA(x = dat, order = c(2, 1, 2)), so here we have taken p=2, d=1, q=2

Coefficients:

	ar1	ar2	ma1	ma2
	1.4733	-0.6877	-1.7620	0.8074
s.e.	0.0735	0.0660	0.0669	0.0627

sigma^2 estimated as 3127: log likelihood = -909.84, aic = 1829.69



For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .

Box-Ljung test

data: e

X-squared = 0.08937, df = 1, p-value = 0.765

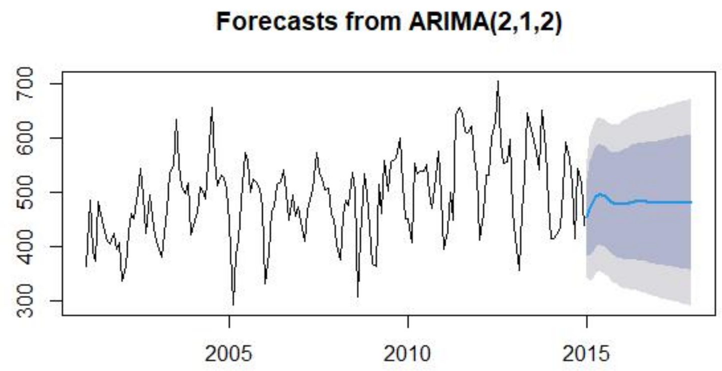
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	453.0148	469.4055	483.2288	492.3235	496.2172	495.6996	492.2597	487.5474
2016	477.4595	478.6724	480.0723	481.3006	482.1477	482.5511	482.5628	482.3027
2017	480.9372	480.9715	481.0681	481.1868	481.2952	481.3733	481.4139	481.4199

	Sep	Oct	Nov	Dec
2015	482.9704	479.4675	477.4541	476.8965
2016	481.9115	481.5139	481.1972	481.0040
2017	481.4009	481.3688	481.3345	481.3060

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

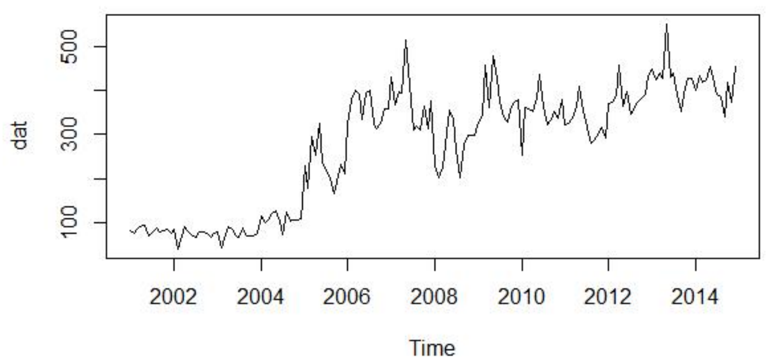
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	362.1887	365.5056	457.4219	474.4045	533.0545	565.2004	572.6451	507.9905
2016	386.6334	374.7726	454.2654	463.3816	518.9851	552.0519	562.9482	502.7458
2017	390.1111	377.0717	455.2612	463.2677	518.1325	550.8741	561.7993	501.8630

	Sep	Oct	Nov	Dec
2015	488.2276	521.1190	494.6824	444.0447
2016	487.1687	523.1654	498.4256	448.1523
2017	486.6581	523.0202	498.5628	448.4543

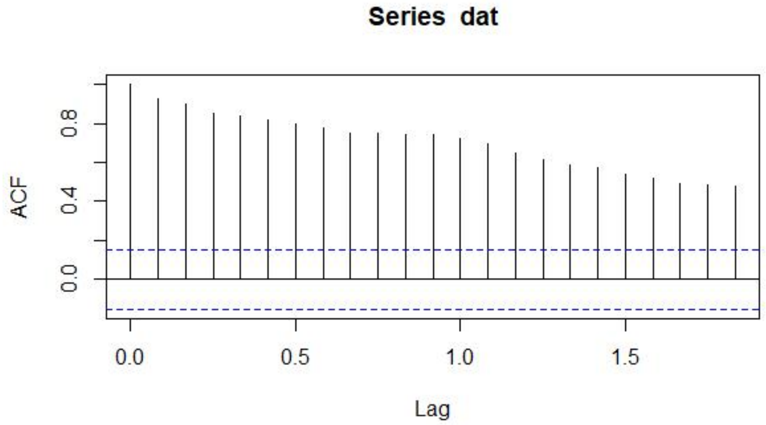
- **<u>Jharkhand :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

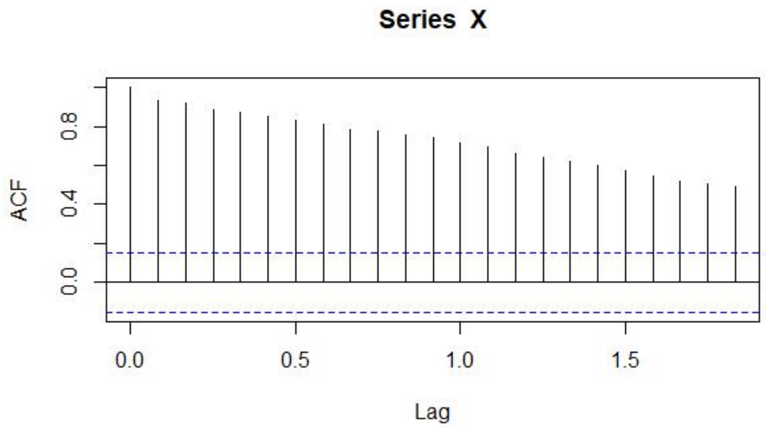


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

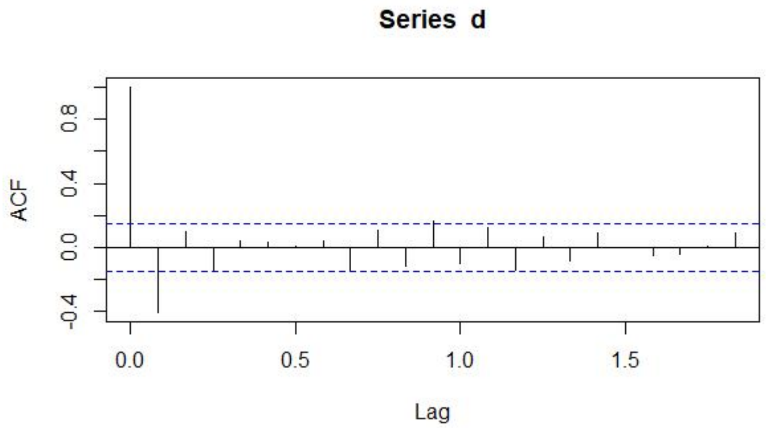


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -2.0262, Lag order = 5, p-value = 0.5654

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

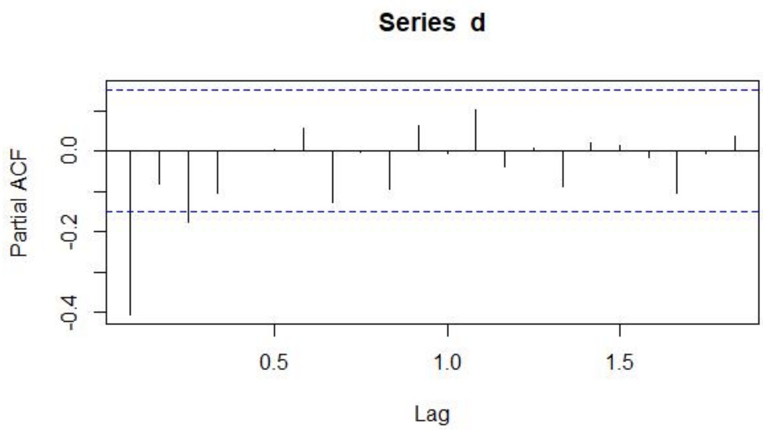
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -6.1283, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

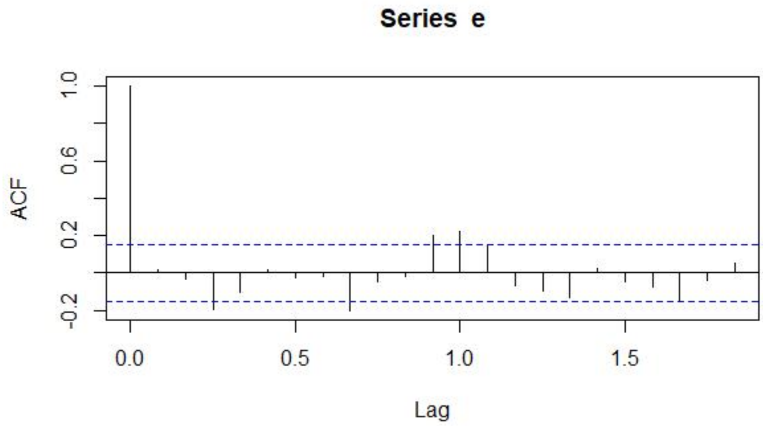
ARIMA(x = dat, order = c(1, 1, 2)), so here we have taken p=1, d=1, q=2

Coefficients:

	ar1	ma1	ma2
	-0.6372	0.2958	-0.1360
s.e.	0.2454	0.2520	0.1387

sigma^2 estimated as 2041: log likelihood = -873.4, aic = 1754.8

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.035392, df = 1, p-value = 0.8508

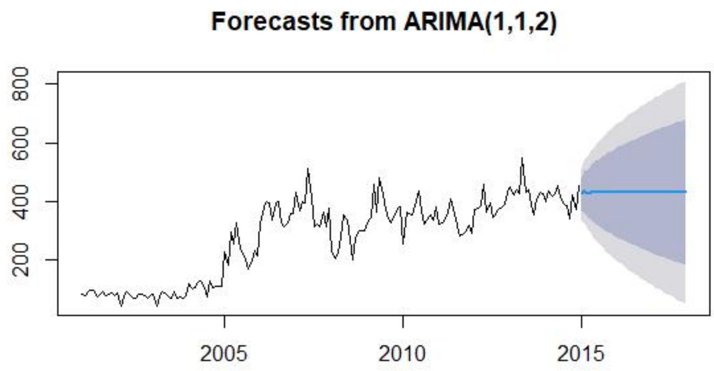
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	424.2864	433.6753	427.6922	431.5049	429.0753	430.6236	429.6369	430.2657
2016	429.9952	430.0373	430.0105	430.0276	430.0167	430.0237	430.0192	430.0221
2017	430.0208	430.0210	430.0209	430.0210	430.0209	430.0210	430.0210	430.0210

	Sep	Oct	Nov	Dec
2015	429.8650	430.1203	429.9576	430.0613
2016	430.0203	430.0214	430.0207	430.0211
2017	430.0210	430.0210	430.0210	430.0210

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

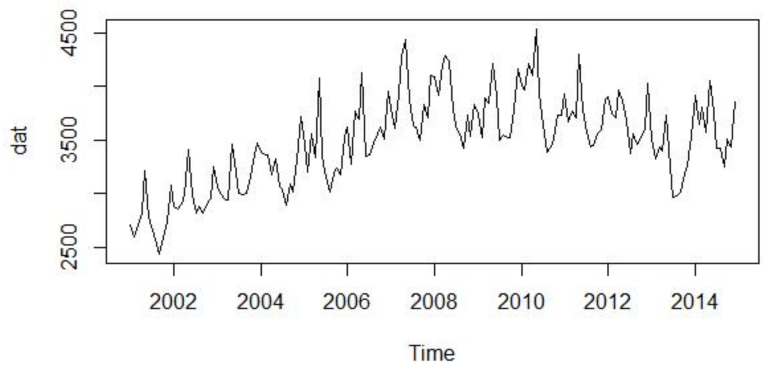
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	422.6651	427.1053	449.4844	454.5247	485.0919	455.0536	418.9131	401.5354
2016	428.3740	423.4674	451.8027	453.0474	486.0333	454.4537	419.2954	401.2918
2017	428.3996	423.4511	451.8131	453.0408	486.0375	454.4510	419.2971	401.2907

	Sep	Oct	Nov	Dec
2015	395.1604	411.4542	414.6184	421.1580
2016	395.3157	411.3553	414.6815	421.1178
2017	395.3164	411.3548	414.6817	421.1176

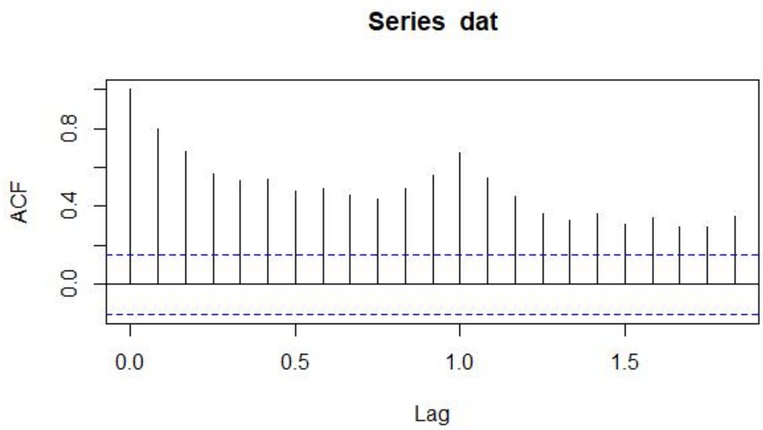
- **Karnataka :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

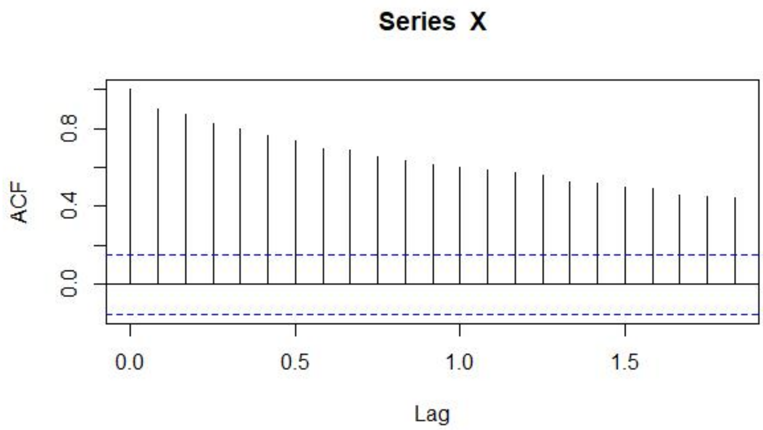


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

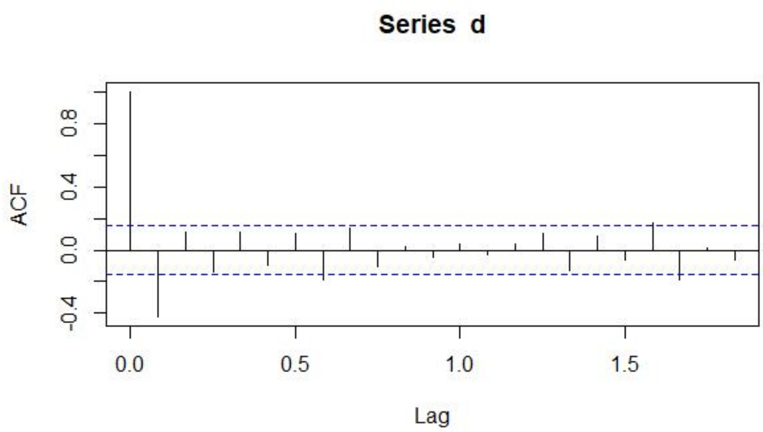


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data’s acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -1.9096, Lag order = 5, p-value = 0.614

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

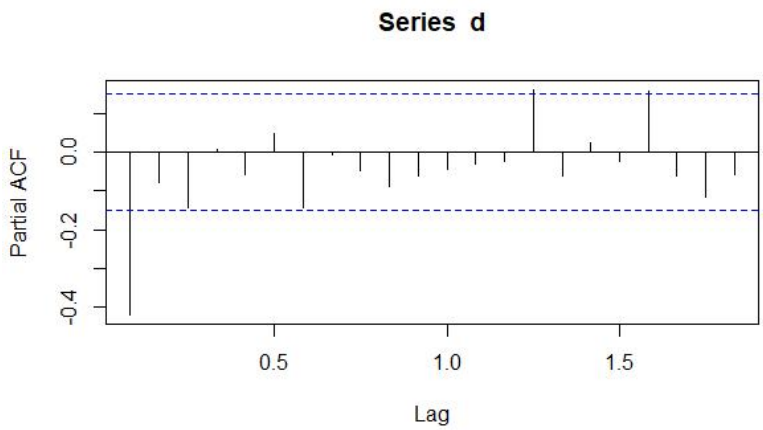
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -5.8855, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

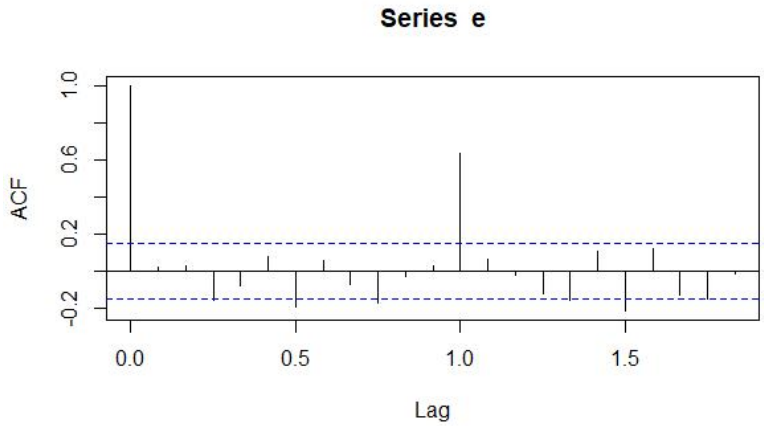
ARIMA(x = dat, order = c(2, 1, 2)), so here we have taken p=2, d=1, q=2

Coefficients:

	ar1	ar2	ma1	ma2
	-0.0167	0.3288	-0.3850	-0.4646
s.e.	0.3486	0.1837	0.3482	0.3091

sigma^2 estimated as 57588: log likelihood = -1152.54, aic = 2315.09

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.10878, df = 1, p-value = 0.7415

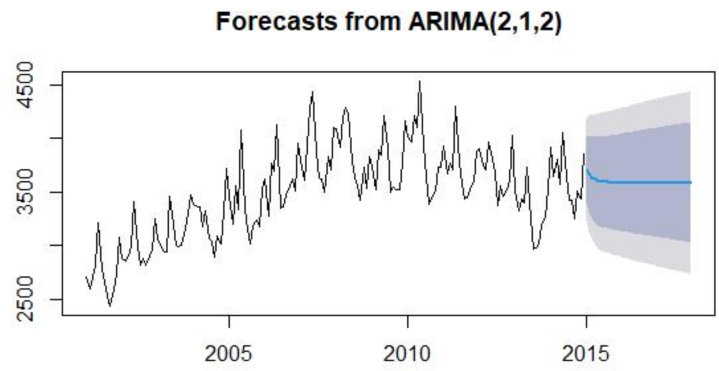
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3711.801	3676.798	3628.328	3617.629	3601.871	3598.616	3593.490	3592.505
2016	3589.731	3589.705	3589.647	3589.639	3589.621	3589.618	3589.612	3589.612
2017	3589.609	3589.608	3589.608	3589.608	3589.608	3589.608	3589.608	3589.608

	Sep	Oct	Nov	Dec
2015	3590.836	3590.540	3589.997	3589.908
2016	3589.610	3589.609	3589.609	3589.609
2017	3589.608	3589.608	3589.608	3589.608

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

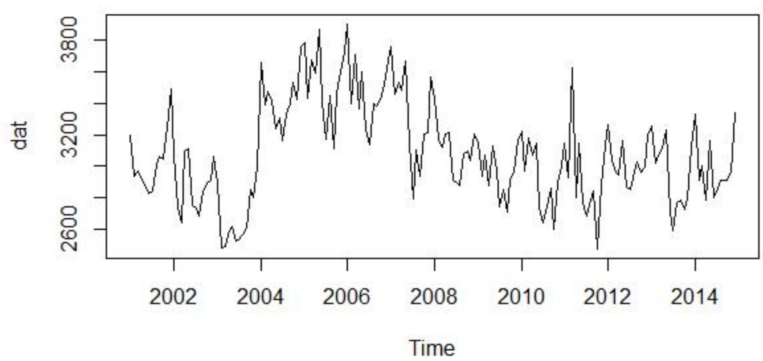
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3842.186	3631.113	3762.944	3730.783	4054.715	3598.175	3377.651	3333.352
2016	3720.116	3544.020	3724.263	3702.794	4042.464	3589.177	3373.773	3330.458
2017	3719.994	3543.923	3724.224	3702.763	4042.452	3589.167	3373.769	3330.455

	Sep	Oct	Nov	Dec
2015	3322.946	3433.528	3495.167	3799.761
2016	3321.719	3432.597	3494.779	3799.462
2017	3321.718	3432.596	3494.779	3799.461

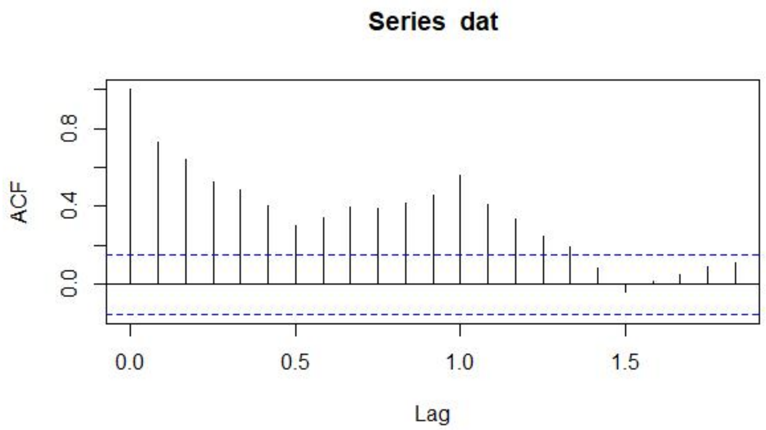
- **Kerela :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

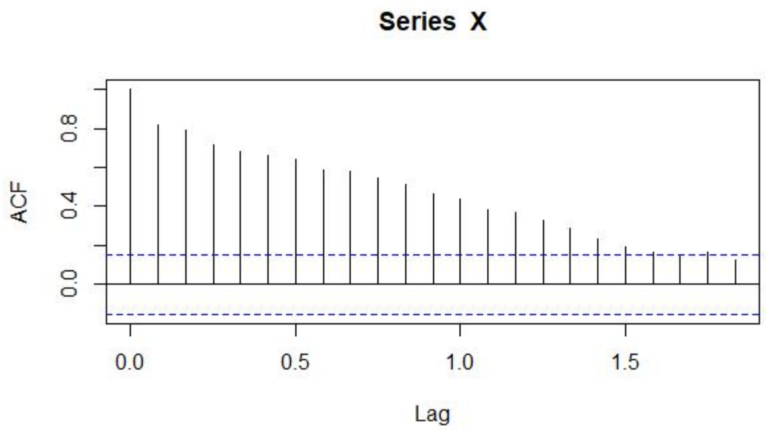


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

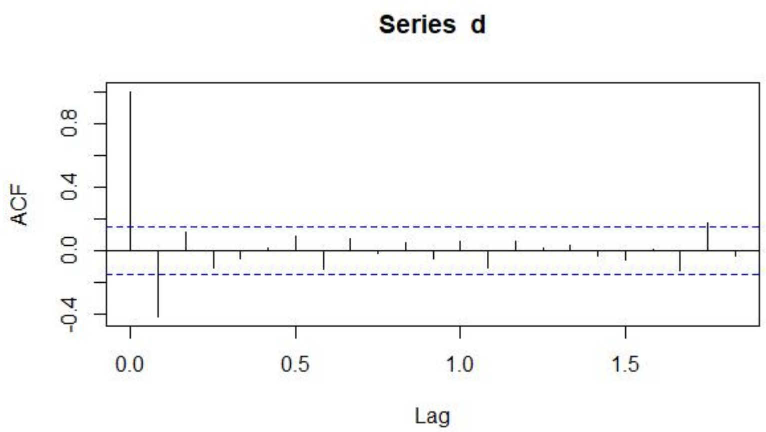


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -1.99, Lag order = 5, p-value = 0.5805

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

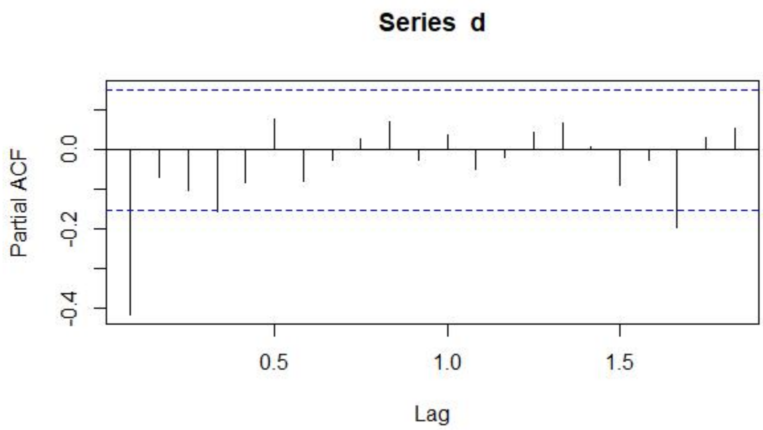
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -6.0587, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

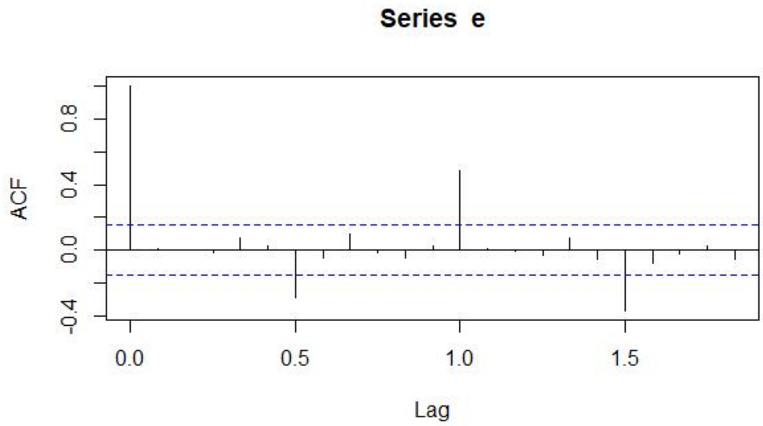
ARIMA(x = dat, order = c(2, 1, 2)), so here we have taken p=2, d=1, q=2

Coefficients:

	ar1	ar2	ma1	ma2
	0.0446	0.3444	-0.4733	-0.3611
s.e.	0.3346	0.1587	0.3395	0.2777

sigma^2 estimated as 44515: log likelihood = -1130.97, aic = 2271.94

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0065483, df = 1, p-value = 0.9355

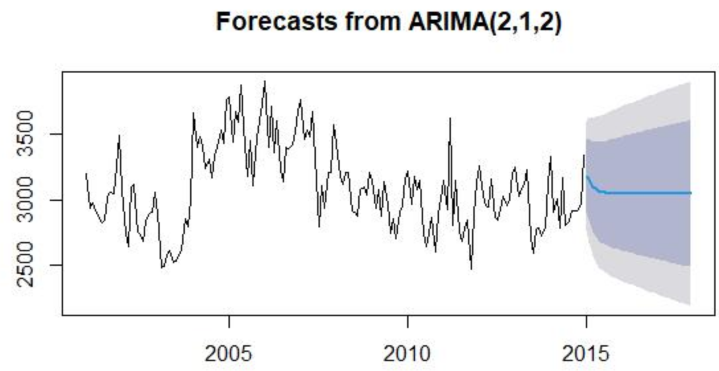
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3172.990	3152.982	3093.878	3084.352	3063.571	3059.363	3052.018	3050.241
2016	3045.346	3045.231	3045.110	3045.065	3045.022	3045.004	3044.989	3044.982
2017	3044.970	3044.970	3044.969	3044.969	3044.969	3044.969	3044.969	3044.969

	Sep	Oct	Nov	Dec
2015	3047.632	3046.903	3045.972	3045.680
2016	3044.976	3044.974	3044.972	3044.971
2017	3044.969	3044.969	3044.969	3044.969

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

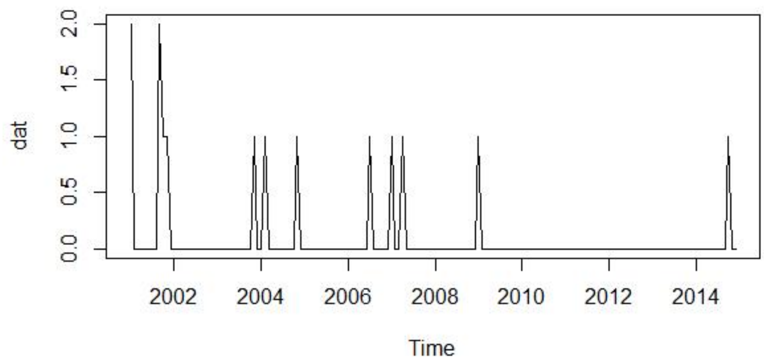
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3451.940	3126.131	3198.592	3092.674	3225.473	2903.429	2789.276	2919.387
2016	3324.297	3018.380	3149.824	3053.388	3206.924	2889.070	2782.247	2914.128
2017	3323.920	3018.119	3149.683	3053.291	3206.872	2889.035	2782.227	2914.115

	Sep	Oct	Nov	Dec
2015	2925.832	2968.328	3048.923	3265.595
2016	2923.177	2966.399	3047.922	3264.886
2017	2923.169	2966.394	3047.919	3264.884

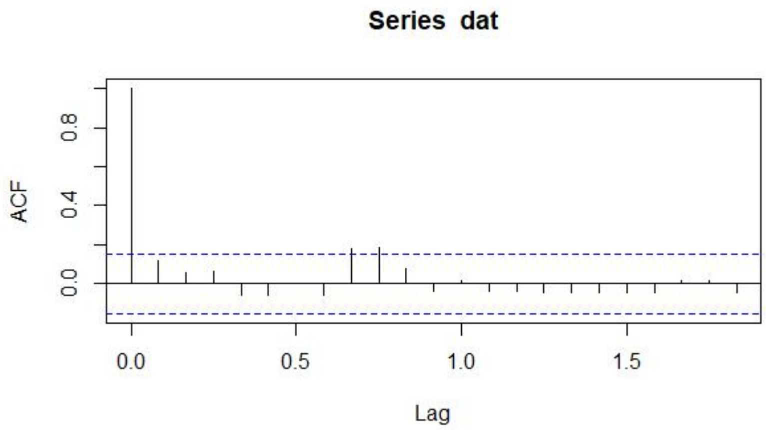
- **Lakshadweep :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is stationary.



So, now we can say that the data is stationary, augmented Dickey Fuller test also showed that

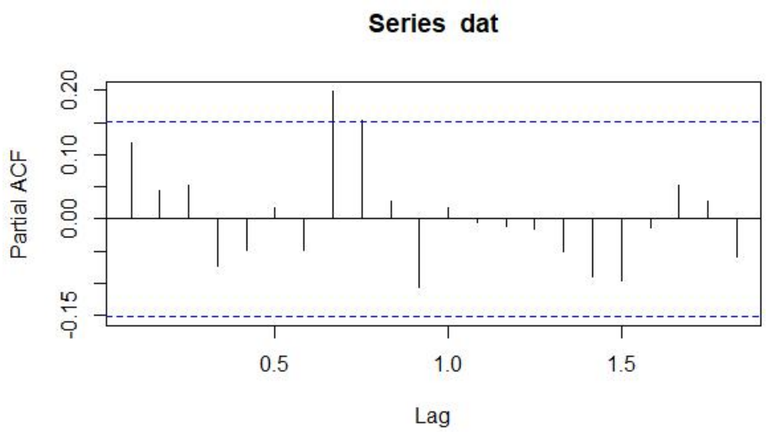
Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -5.6499, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the data:



Therefore the data is stationary at d=0.

Fitted model by checking AIC is:

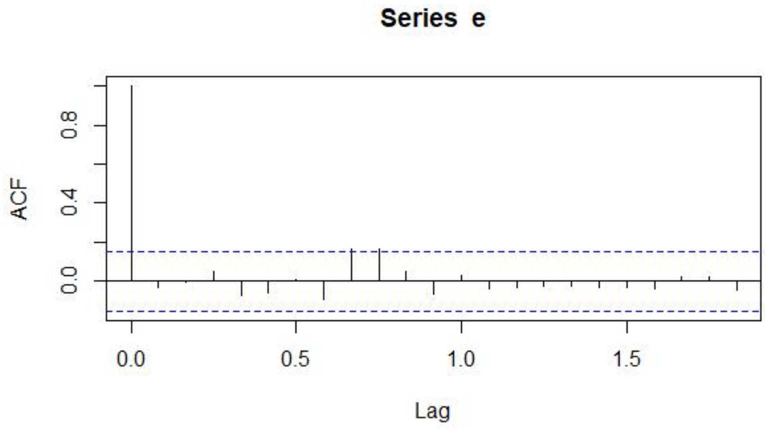
ARIMA(X,order = c(1, 0, 1)), so here we have taken p=1, d=0, q=1

Coefficients:

	ar1	ma1	intercept
	0.4479	-0.3035	0.0862
s.e.	0.3672	0.3898	0.0305

sigma^2 estimated as 0.09816: log likelihood = -43.42, aic = 94.83

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.19091, df = 1, p-value = 0.6622

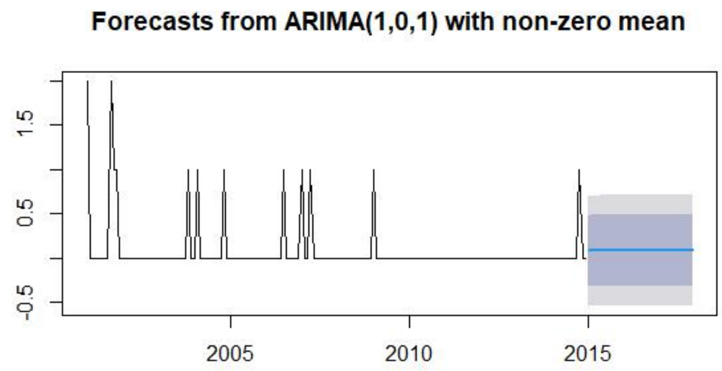
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul
2015	0.08163566	0.08415905	0.08528927	0.08579548	0.08602221	0.08612376	0.08616925
2016	0.08620585	0.08620602	0.08620609	0.08620612	0.08620614	0.08620614	0.08620615
2017	0.08620615	0.08620615	0.08620615	0.08620615	0.08620615	0.08620615	0.08620615

	Aug	Sep	Oct	Nov	Dec
2015	0.08618962	0.08619875	0.08620283	0.08620466	0.08620548
2016	0.08620615	0.08620615	0.08620615	0.08620615	0.08620615
2017	0.08620615	0.08620615	0.08620615	0.08620615	0.08620615

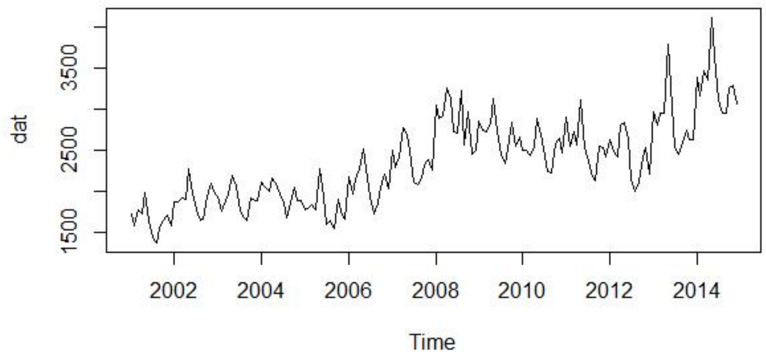
Plot for forecast is given below:



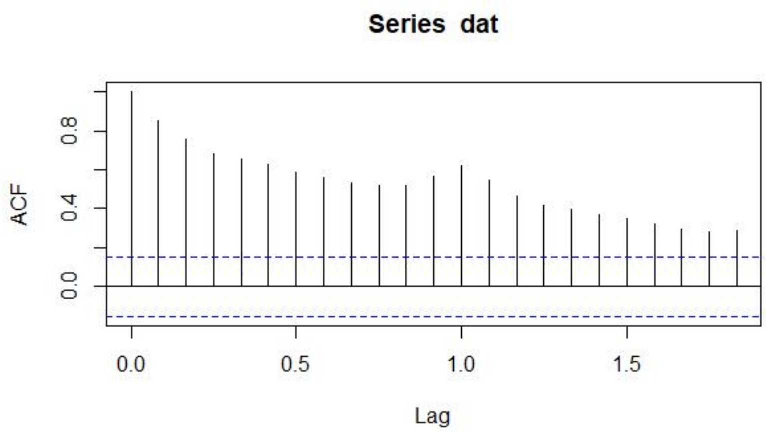
- **<u>Madhya Pradesh :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

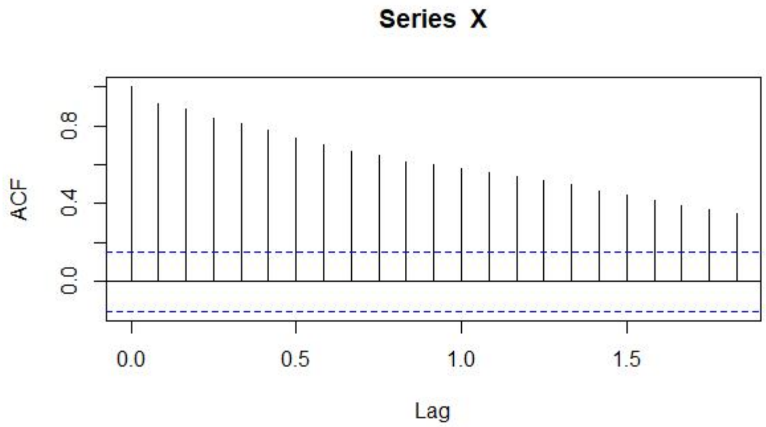


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

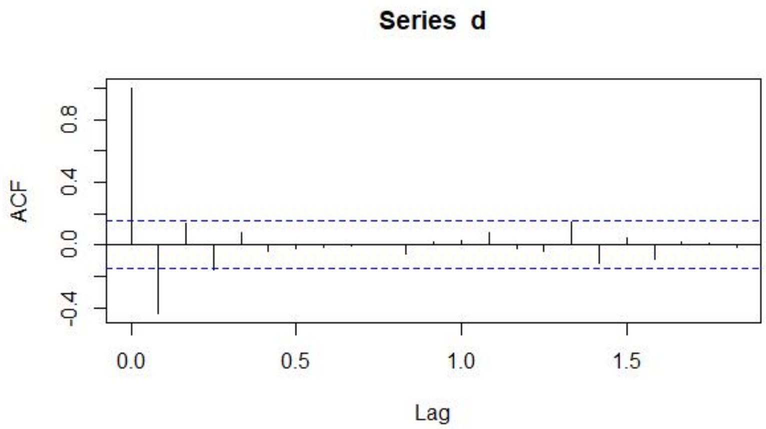


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -2.5423, Lag order = 5, p-value = 0.3499

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

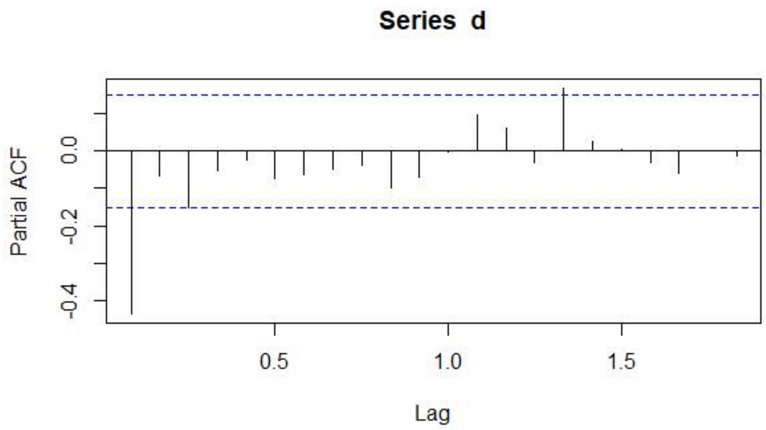
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -6.503, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

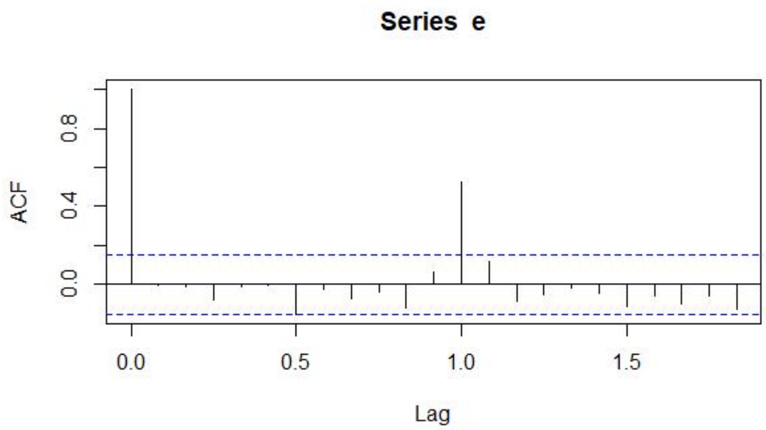
ARIMA(x = dat, order = c(2, 1, 2)), so here we have taken p=2, d=1, q=2

Coefficients:

	ar1	ar2	ma1	ma2
	-0.4765	0.5224	0.1251	-0.8634
s.e.	0.0925	0.0923	0.0523	0.0511

sigma^2 estimated as 63714: log likelihood = -1161.36, aic = 2332.72

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0052645, df = 1, p-value = 0.9422

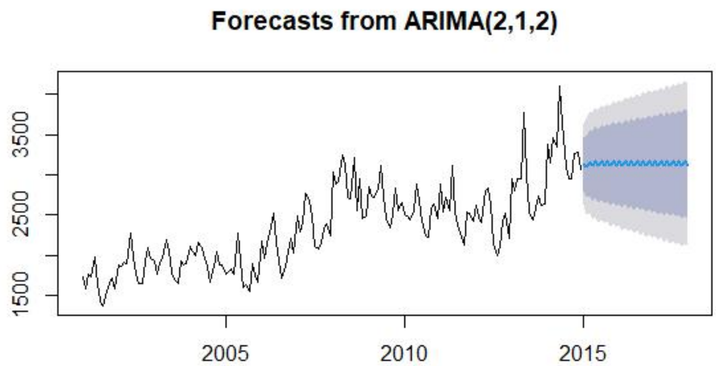
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3132.596	3098.868	3152.865	3109.517	3158.381	3112.452	3159.865	3113.278
2016	3160.324	3113.671	3160.299	3113.708	3160.268	3113.743	3160.235	3113.776
2017	3160.136	3113.876	3160.103	3113.909	3160.070	3113.942	3160.037	3113.974

	Sep	Oct	Nov	Dec
2015	3160.246	3113.528	3160.326	3113.621
2016	3160.202	3113.809	3160.169	3113.843
2017	3160.004	3114.007	3159.971	3114.040

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

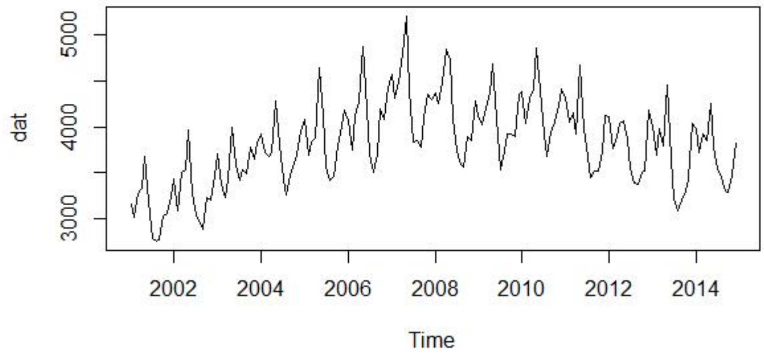
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3290.820	3112.275	3233.641	3280.635	3618.359	3219.849	2966.874	2845.894
2016	3318.548	3127.078	3241.074	3284.827	3620.245	3221.140	2967.245	2846.392
2017	3318.360	3127.283	3240.879	3285.027	3620.048	3221.339	2967.047	2846.590

	Sep	Oct	Nov	Dec
2015	2890.208	3101.916	3084.714	2940.358
2016	2890.164	3102.197	3084.557	2940.580
2017	2889.966	3102.395	3084.359	2940.777

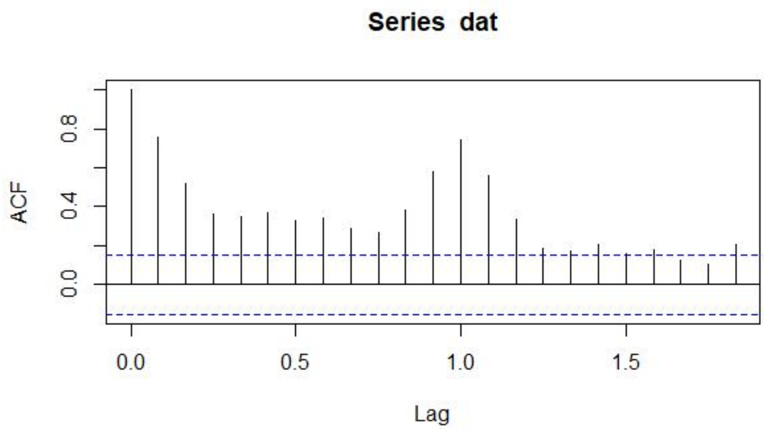
➢ **Maharashtra:**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

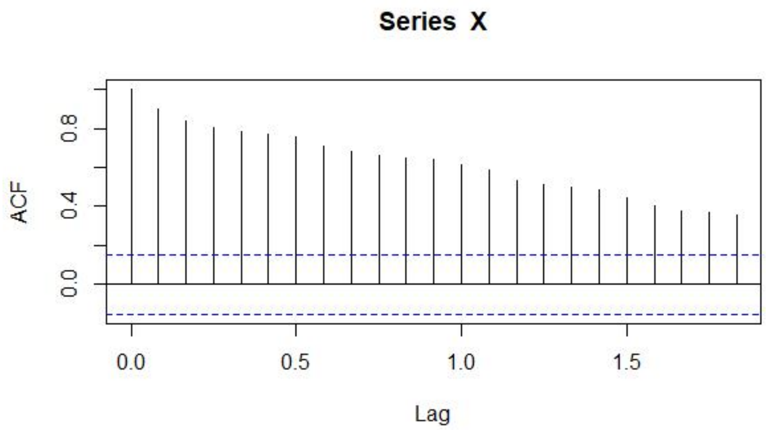


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

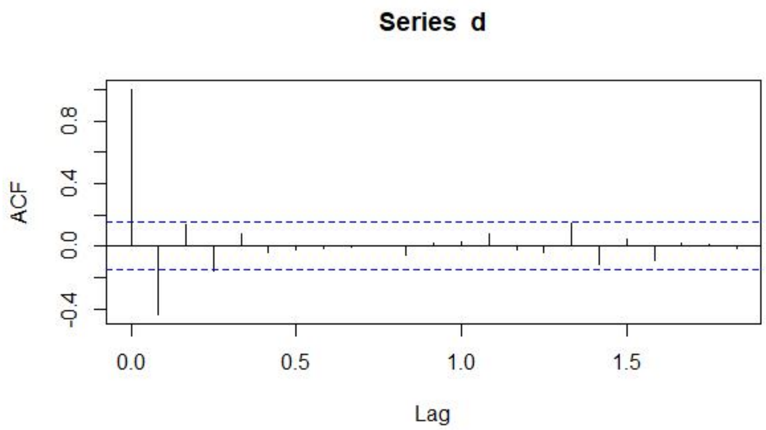


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -1.3973, Lag order = 5, p-value = 0.8279

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

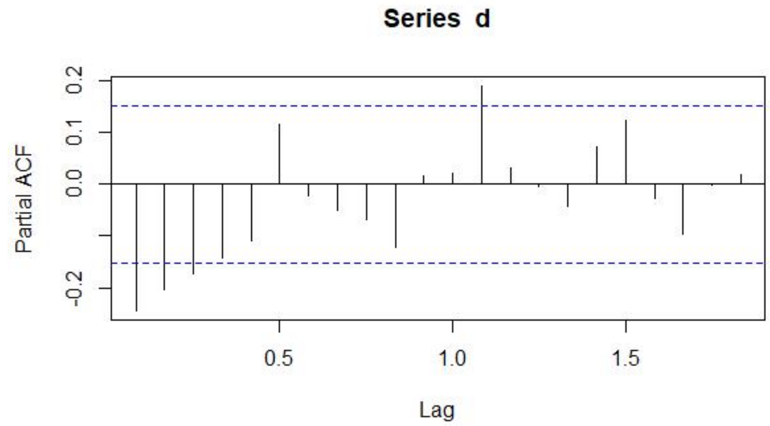
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -6.4138, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

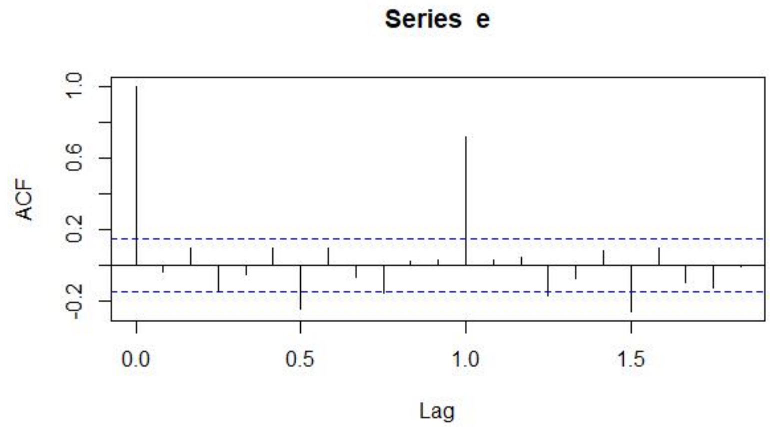
ARIMA(x = dat, order = c(2, 1, 1)), so here we have taken p=2, d=1, q=1

Coefficients:

	ar1	ar2	ma1
	0.6999	-0.2508	-0.8920
s.e.	0.0796	0.0767	0.0336

sigma^2 estimated as 81757: log likelihood = -1181.97, aic = 2371.94

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.28384, df = 1, p-value = 0.5942

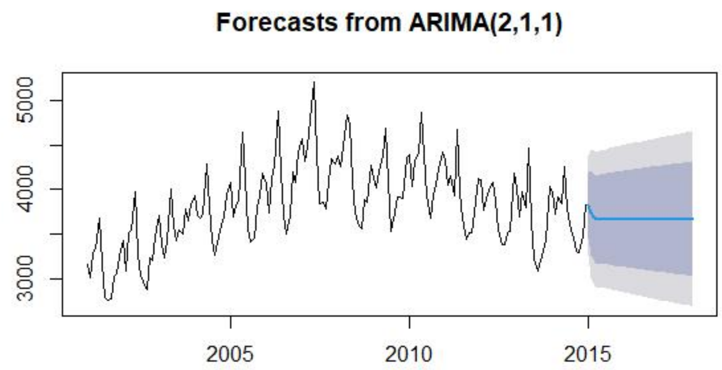
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3819.655	3735.540	3676.256	3655.862	3656.457	3661.989	3665.711	3666.929
2016	3666.187	3666.211	3666.226	3666.230	3666.229	3666.228	3666.227	3666.227
2017	3666.227	3666.227	3666.227	3666.227	3666.227	3666.227	3666.227	3666.227

	Sep	Oct	Nov	Dec
2015	3666.847	3666.485	3666.252	3666.180
2016	3666.227	3666.227	3666.227	3666.227
2017	3666.227	3666.227	3666.227	3666.227

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

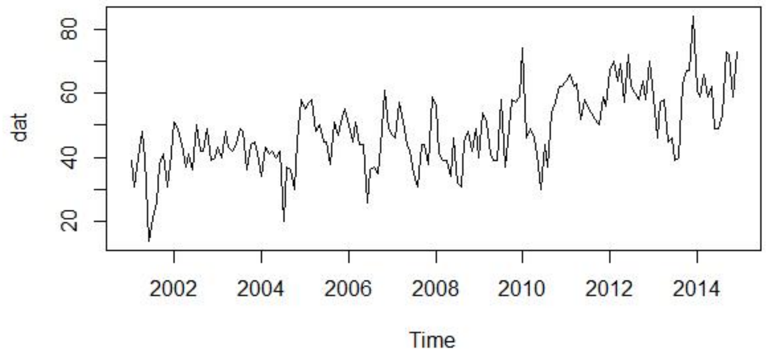
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	4047.092	3679.836	3796.561	3853.779	4304.554	3732.221	3326.792	3229.385
2016	3893.624	3610.507	3786.531	3864.147	4314.326	3736.459	3327.308	3228.682
2017	3893.663	3610.522	3786.532	3864.144	4314.324	3736.458	3327.308	3228.683

	Sep	Oct	Nov	Dec
2015	3294.678	3505.819	3559.362	3874.084
2016	3294.057	3505.561	3559.336	3874.131
2017	3294.058	3505.561	3559.336	3874.131

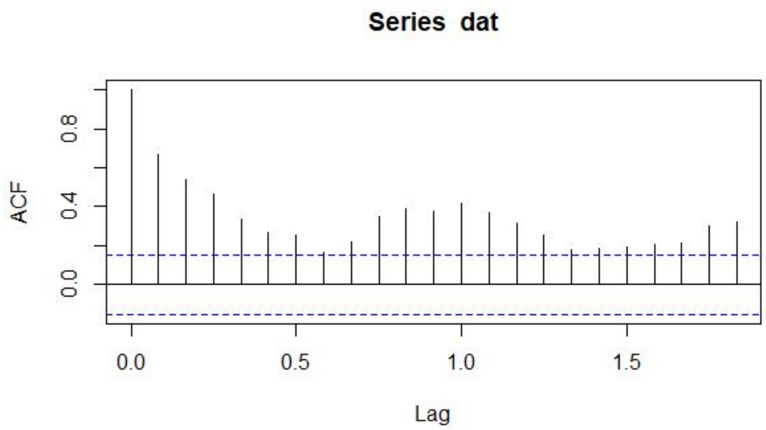
- **Manipur :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

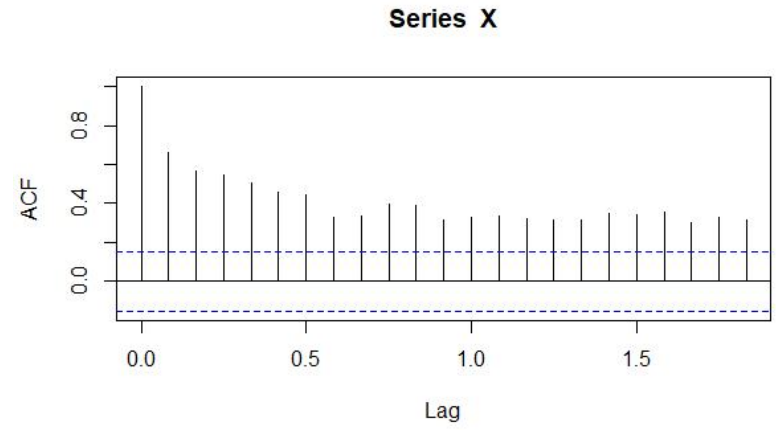


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.



So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



So, now we can say that the data is stationary, augmented Dickey Fuller test also showed that

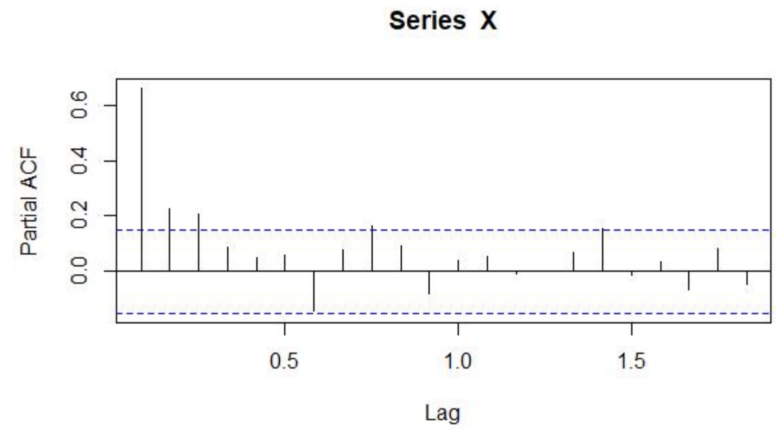
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -3.9231, Lag order = 5, p-value = 0.01464

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=0.

Fitted model by checking AIC is:

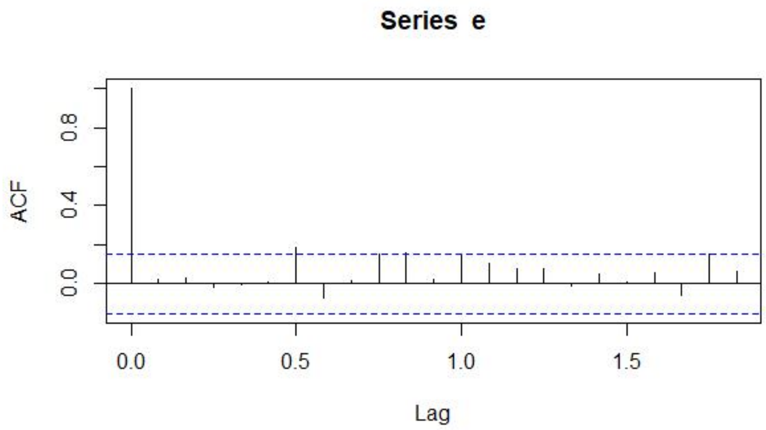
ARIMA(x = dat, order = c(2, 0, 3)), so here we have taken p=2, d=0, q=3

Coefficients:

	ar1	ar2	ma1	ma2	ma3	intercept
	1.2994	-0.5580	-0.7680	0.2587	0.2885	48.4891
s.e.	0.1831	0.1793	0.1885	0.1171	0.0872	1.9112

sigma^2 estimated as 68.67: log likelihood = -594.2, aic = 1202.4

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.071028, df = 1, p-value = 0.7898

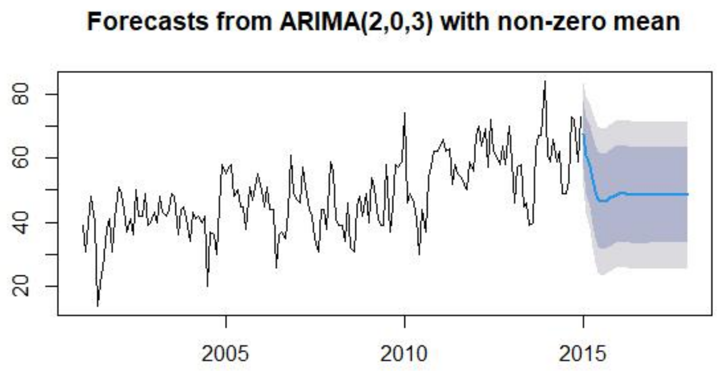
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	67.30763	60.98552	57.59307	53.34539	49.71902	47.37724	46.35798	46.34034
2016	48.84294	48.85776	48.77067	48.64924	48.54006	48.46594	48.43057	48.42596
2017	48.49870	48.49986	48.49771	48.49428	48.49102	48.48870	48.48750	48.48724

	Sep	Oct	Nov	Dec
2015	46.88622	47.60539	48.23526	48.65240
2016	48.43971	48.46015	48.47904	48.49217
2017	48.48757	48.48814	48.48870	48.48911

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

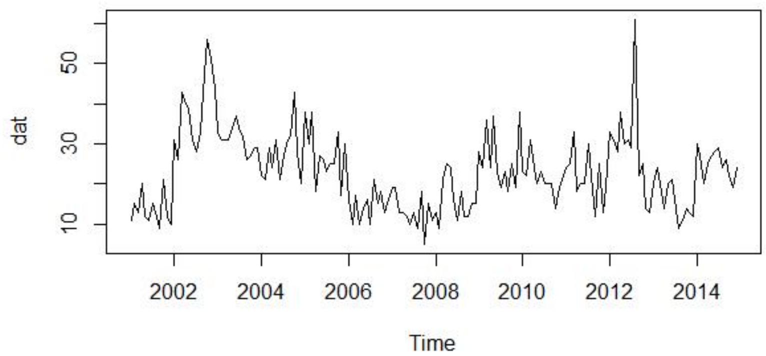
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	72.85972	63.36132	61.69323	53.61863	46.03392	42.26907	40.92929	38.59755
2016	54.39503	51.23356	52.87083	48.92248	44.85496	43.35777	43.00188	40.68317
2017	54.05079	50.87566	52.59787	48.76752	44.80592	43.38052	43.05881	40.74445

	Sep	Oct	Nov	Dec
2015	45.20112	49.10939	50.63350	56.09872
2016	46.75461	49.96416	50.87728	55.93849
2017	46.80247	49.99215	50.88694	55.93543

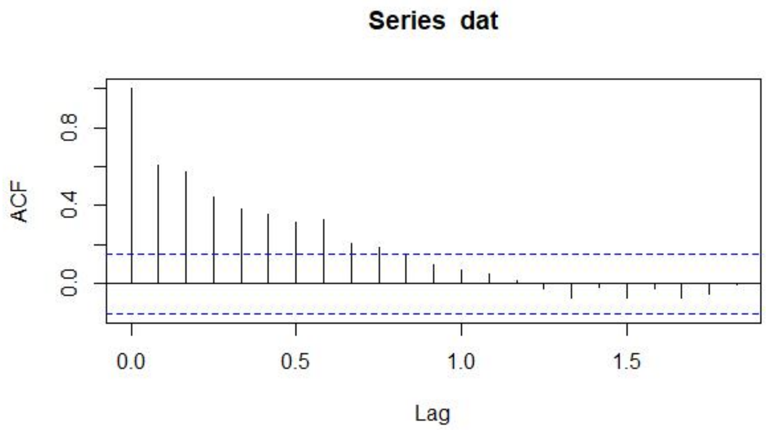
- **Meghalaya :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

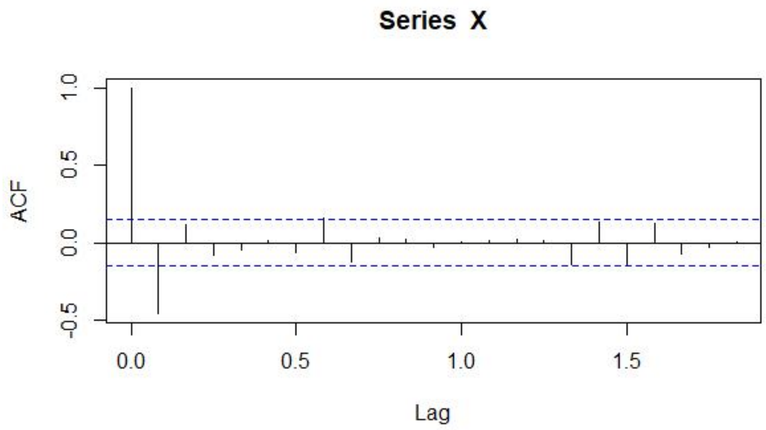
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -3.2337, Lag order = 5, p-value = 0.08457

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

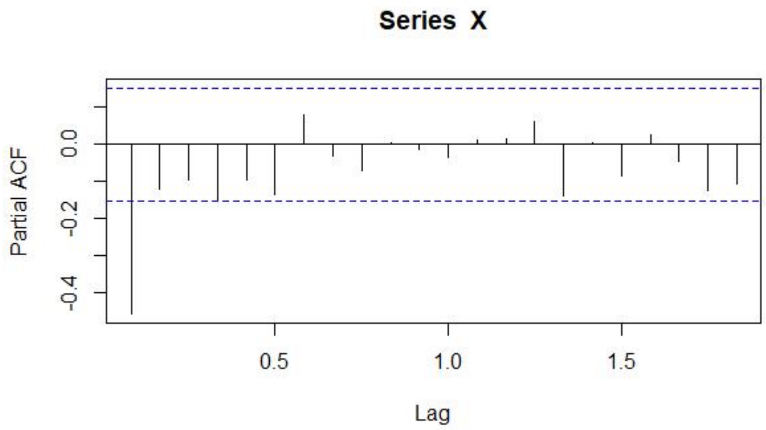
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -7.5944, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

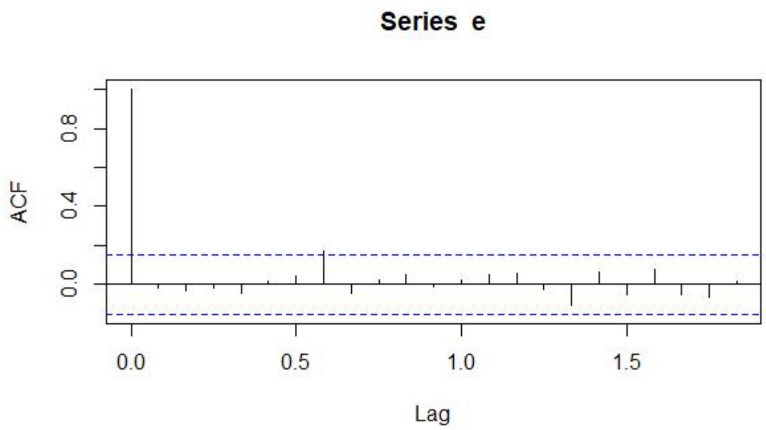
ARIMA(X,order = c (2, 1, 1)), so here we have taken p=2, d=1, q=1

Coefficients:

	ar1	ar2	ma1
	0.4183	0.3319	-1.0000
s.e.	0.0733	0.0735	0.0205

sigma^2 estimated as 51.36: log likelihood = -567.43, aic = 1142.86

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.032883, df = 1, p-value = 0.8561

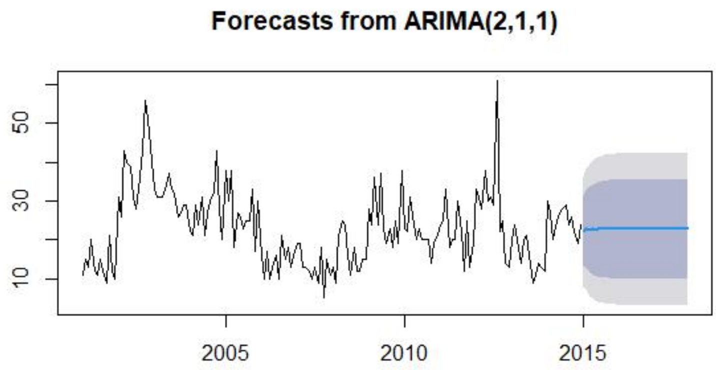
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	22.01989	22.85131	22.54182	22.68834	22.64689	22.67819	22.67752	22.68763
2016	22.70565	22.70771	22.70937	22.71075	22.71188	22.71281	22.71358	22.71421
2017	22.71602	22.71621	22.71637	22.71650	22.71661	22.71670	22.71677	22.71683

	Sep	Oct	Nov	Dec
2015	22.69164	22.69667	22.70011	22.70321
2016	22.71472	22.71515	22.71550	22.71578
2017	22.71688	22.71692	22.71695	22.7169

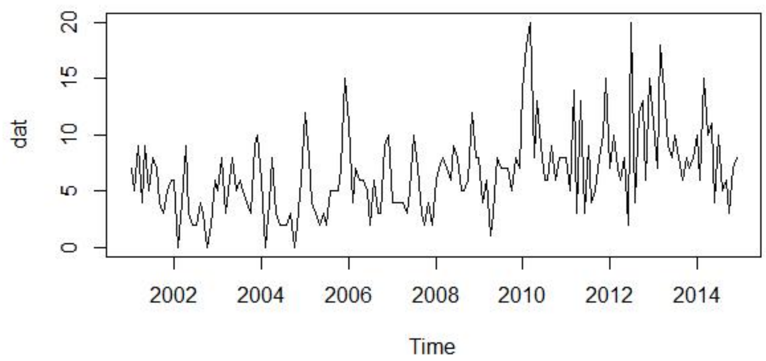
Plot for forecast is given below:



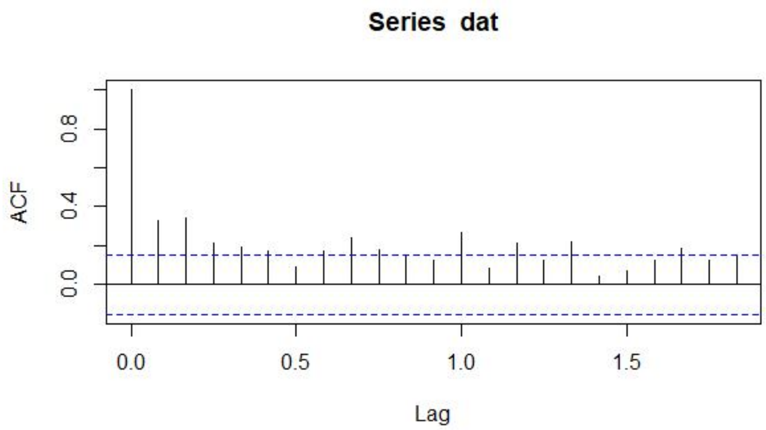
- **<u>Mizoram :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is stationary.



So, now we can say that the data is stationary, augmented Dickey Fuller test also showed that

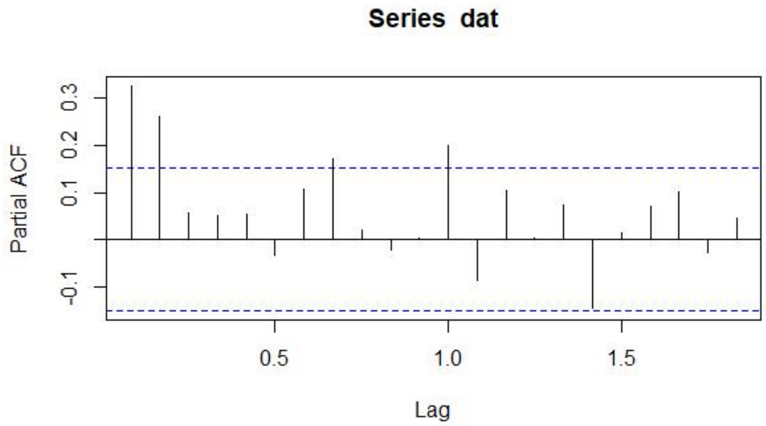
Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -5.479, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=0.

Fitted model by checking AIC is:

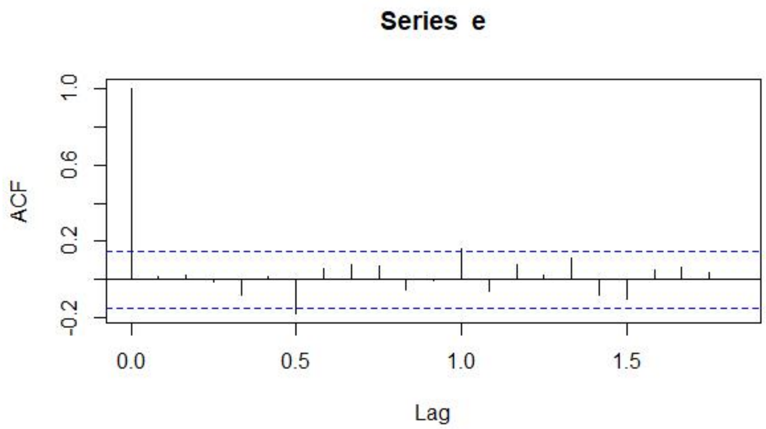
ARIMA(X,order = c (2, 0, 2)), so here we have taken p=2, d=0, q=2

Coefficients:

	ar1	ar2	ma1	ma2	intercept
	-0.1363	0.8547	0.3586	-0.6036	6.6560
s.e.	0.1047	0.1047	0.1760	0.1758	0.6845

sigma^2 estimated as 11.53: log likelihood = -444.2, aic = 900.4

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.060826, df = 1, p-value = 0.8052

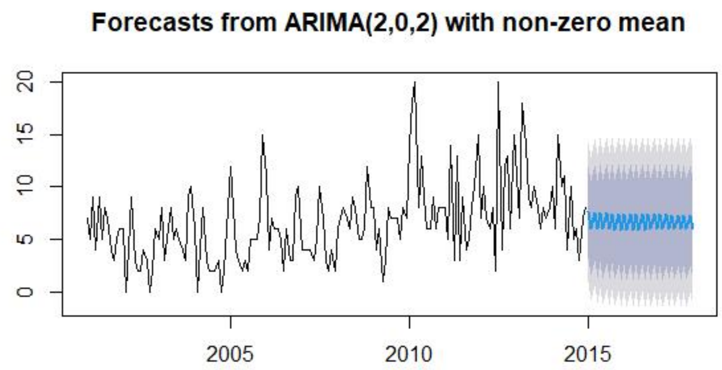
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	7.614561	6.037147	7.559558	6.003962	7.517070	5.981390	7.483833	5.966627
2016	7.418533	5.951091	7.403745	5.951643	7.391032	5.953848	7.379865	5.957254
2017	7.352247	5.971767	7.344276	5.977417	7.336693	5.983280	7.329414	5.989282

	Sep	Oct	Nov	Dec
2015	7.457438	5.957605	7.436108	5.952802
2016	7.369858	5.961528	7.360722	5.966426
2017	7.322375	5.995372	7.315529	6.001509

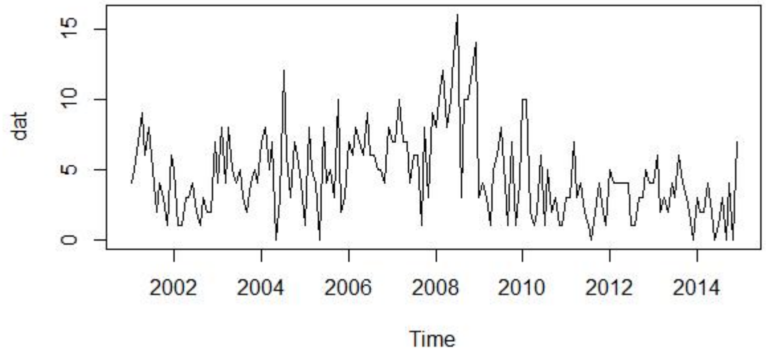
Plot for forecast is given below:



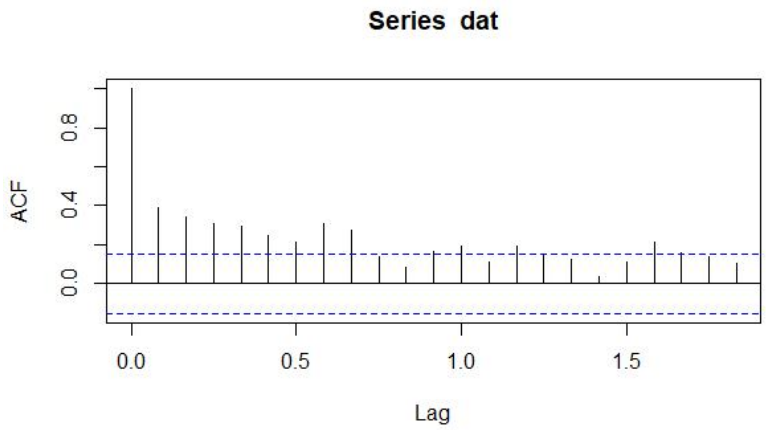
- **<u>Nagaland :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

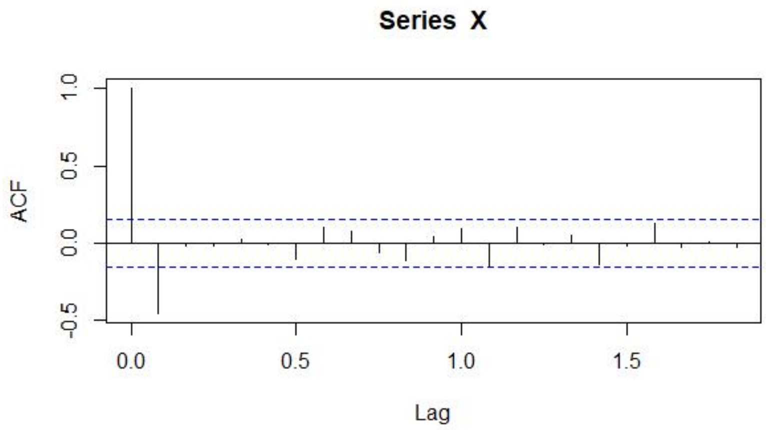
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -3.2104, Lag order = 5, p-value = 0.08845

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

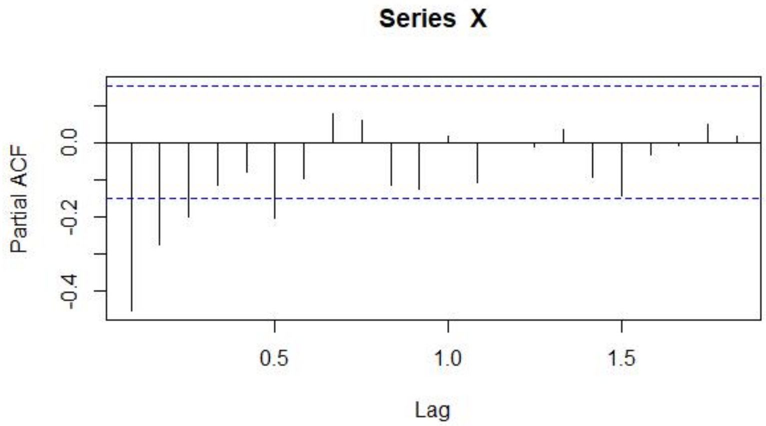
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -8.6977, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

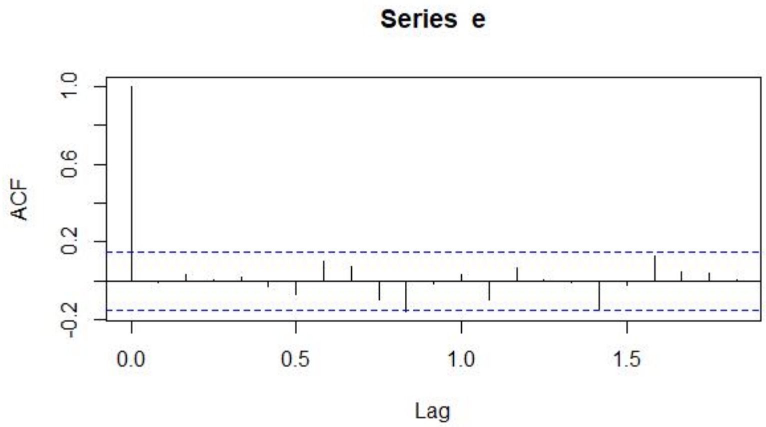
ARIMA(X,order = c (1, 1, 1)), so here we have taken p=1, d=1, q=1

Coefficients:

	ar1	ma1
	0.0997	-0.8444
s.e.	0.1035	0.0633

sigma^2 estimated as 7.325: log likelihood = -403.78, aic = 813.56

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0098186, df = 1, p-value = 0.9211

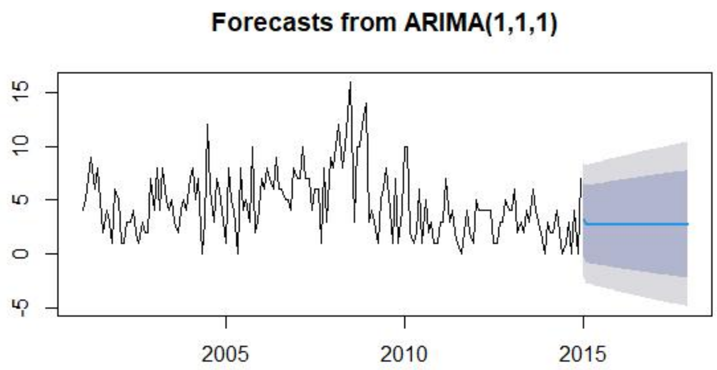
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	3.178210	2.797102	2.759097	2.755308	2.754930	2.754892	2.754888	2.754888
2016	2.754888	2.754888	2.754888	2.754888	2.754888	2.754888	2.754888	2.754888
2017	2.754888	2.754888	2.754888	2.754888	2.754888	2.754888	2.754888	2.754888

	Sep	Oct	Nov	Dec
2015	2.754888	2.754888	2.754888	2.754888
2016	2.754888	2.754888	2.754888	2.754888
2017	2.754888	2.754888	2.754888	2.754888

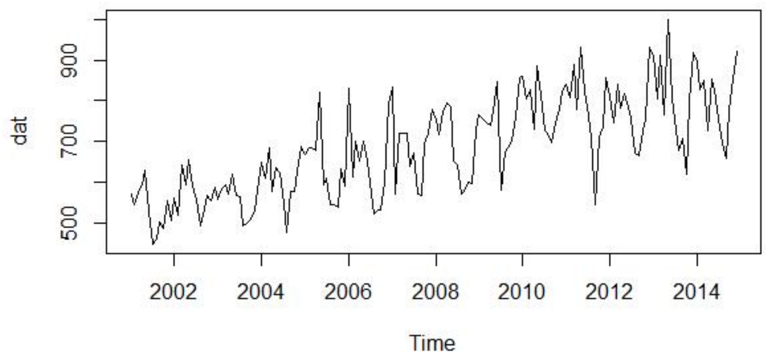
Plot for forecast is given below:



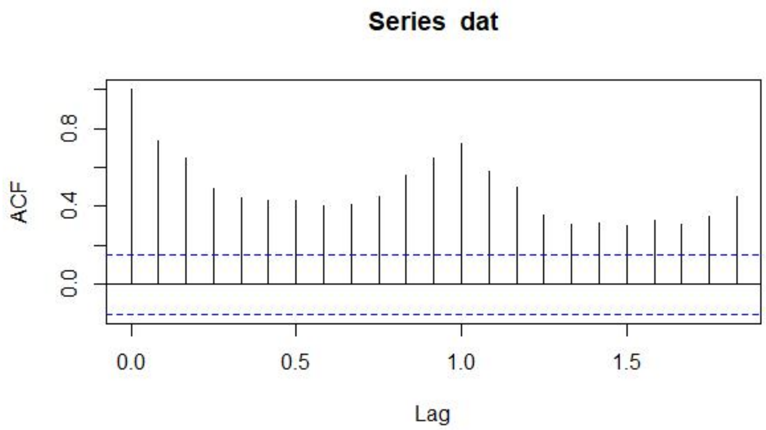
➢ **Odisha :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

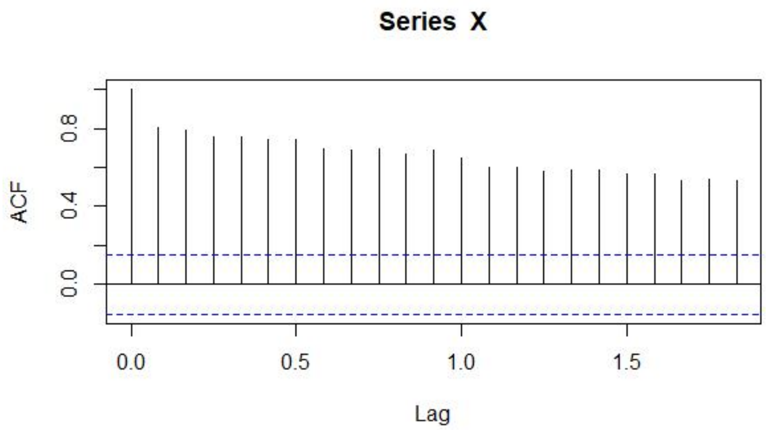


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.



So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



So, now we can say that the data is stationary, augmented Dickey Fuller test also showed that

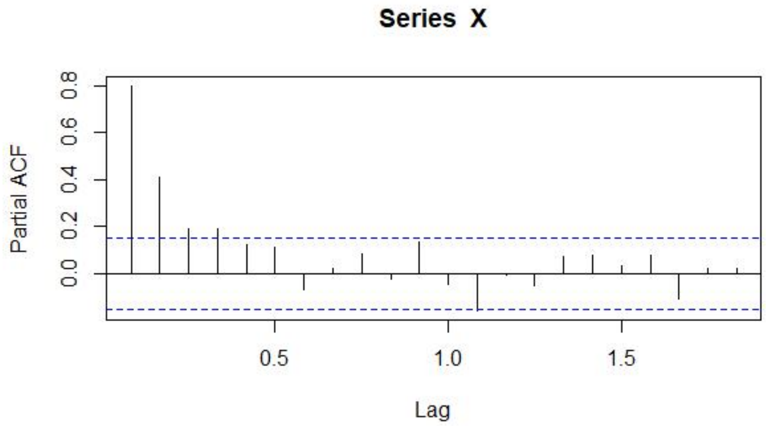
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -3.5564, Lag order = 5, p-value = 0.03936

alternative hypothesis: stationary

The pacf plot for the data:



Therefore the data is stationary at d=0.

Fitted model by checking AIC is:

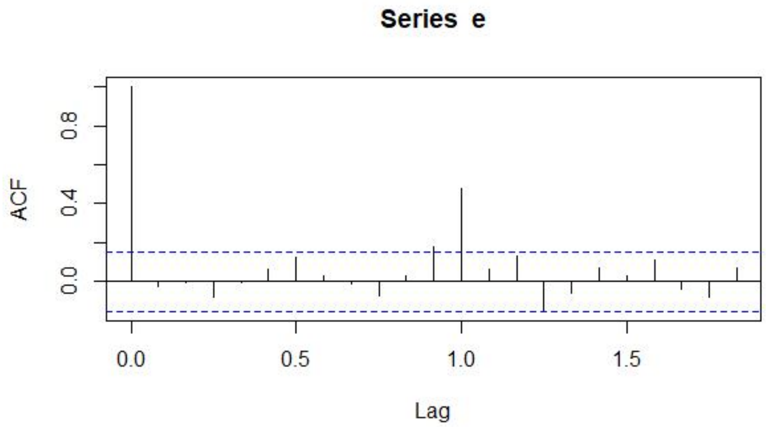
ARIMA(x = dat, order = c(1, 0, 2)), so here we have taken p=1, d=0, q=2

Coefficients:

	ar1	ma1	ma2	intercept
	0.7160	-0.0674	0.2530	688.4224
s.e.	0.1036	0.1253	0.1183	24.3239

sigma^2 estimated as 5894: log likelihood = -968.15, aic = 1946.29

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.082053, df = 1, p-value = 0.7745

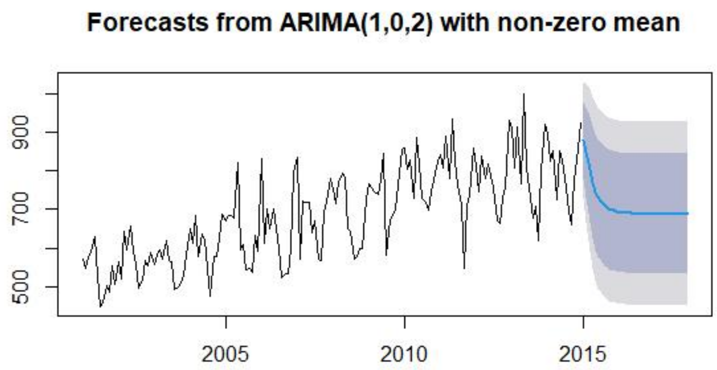
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	878.8836	848.0946	802.7470	770.2784	747.0309	730.3859	718.4681	709.9350
2016	692.4704	691.3208	690.4976	689.9082	689.4862	689.1841	688.9678	688.8129
2017	688.4959	688.4750	688.4600	688.4493	688.4417	688.4362	688.4323	688.4295

	Sep	Oct	Nov	Dec
2015	703.8253	699.4508	696.3187	694.0761
2016	688.7020	688.6226	688.5657	688.5250
2017	688.4274	688.4260	688.4250	688.4242

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

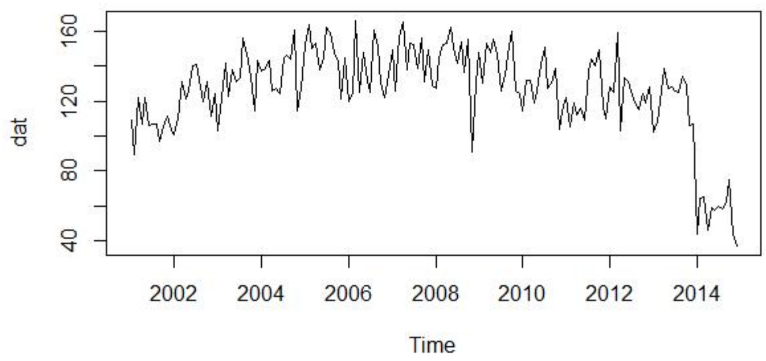
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	956.2463	854.5727	870.9719	778.9904	837.6244	744.0018	673.8372	613.7560
2016	769.8331	697.7989	758.7225	698.6203	780.0797	702.8000	644.3369	592.6339
2017	765.8586	694.9531	756.6849	697.1614	779.0352	702.0521	643.8014	592.2505

	Sep	Oct	Nov	Dec
2015	608.7874	633.1020	677.5276	750.0766
2016	593.6640	622.2737	669.7746	744.5255
2017	593.3895	622.0772	669.6338	744.4248

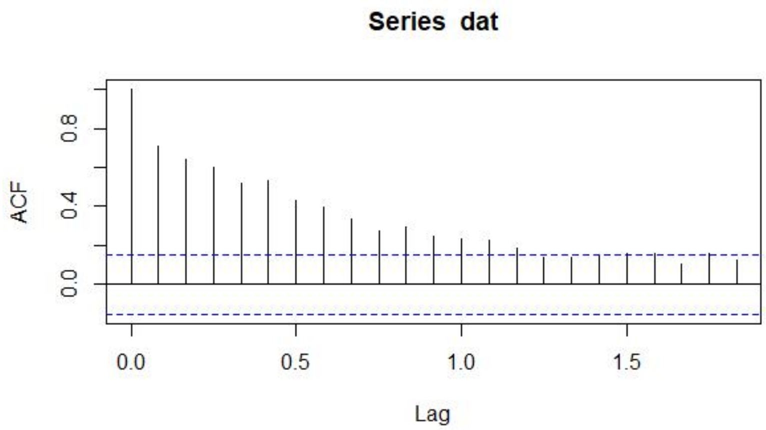
➤ **Puducherry :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

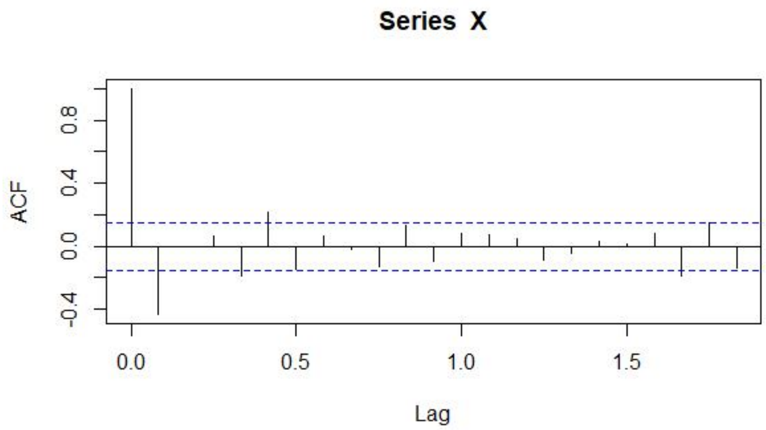
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -1.3134, Lag order = 5, p-value = 0.8629

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

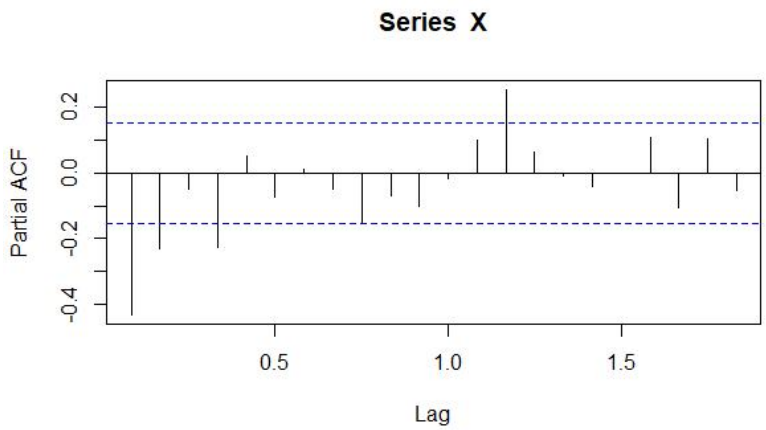
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -7.0557, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

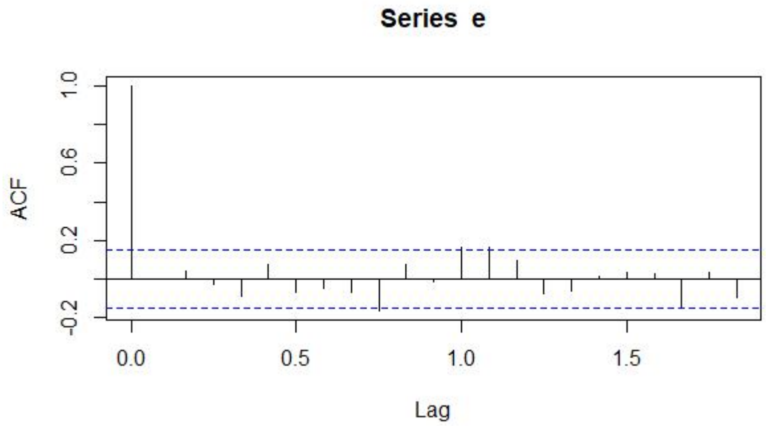
ARIMA(X,order = c (2, 1, 2)), so here we have taken p=2, d=1, q=2

Coefficients:

	ar1	ar2	ma1	ma2
	-0.9356	-0.0281	0.3870	-0.5837
s.e.	0.1526	0.1533	0.1351	0.1356

sigma^2 estimated as 242.8: log likelihood = -696.08, aic = 1402.17

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0020262, df = 1, p-value = 0.9641

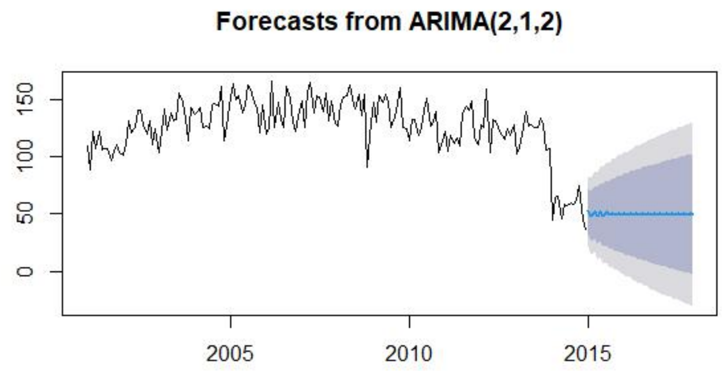
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	52.36104	47.78199	51.63504	48.15854	51.30311	48.45854	51.03174	48.70402
2016	50.47926	49.20379	50.35758	49.31387	50.25801	49.40394	50.17653	49.47764
2017	50.01065	49.62770	49.97412	49.66075	49.94422	49.68779	49.91976	49.70992

	Sep	Oct	Nov	Dec
2015	50.80967	48.90490	50.62796	49.06928
2016	50.10986	49.53796	50.05530	49.58731
2017	49.89974	49.72803	49.88336	49.74285

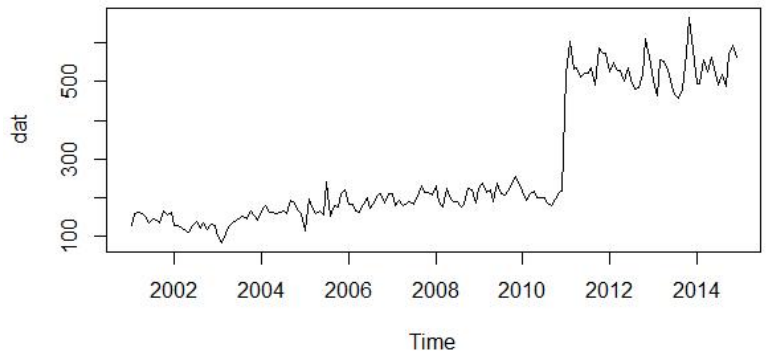
Plot for forecast is given below:



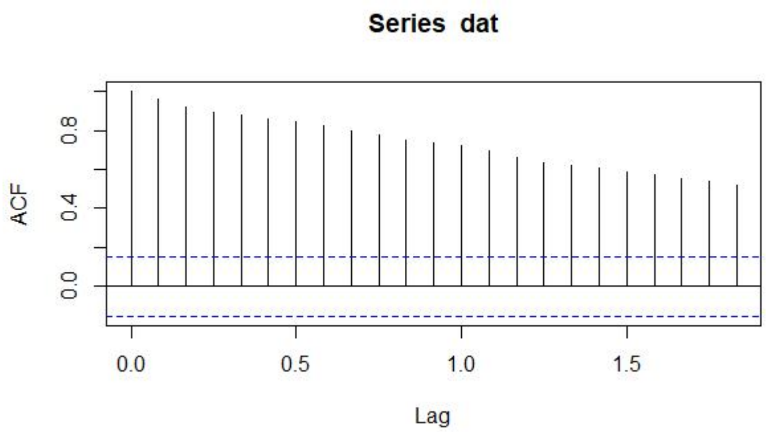
➢ **<u>Punjab :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

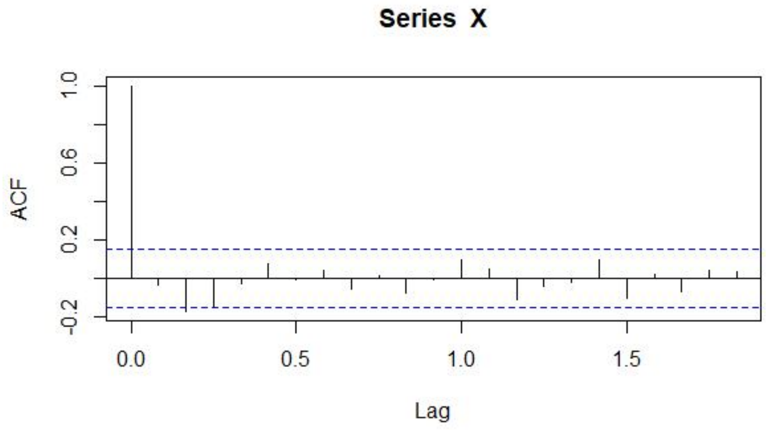
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -2.2374, Lag order = 5, p-value = 0.4772

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

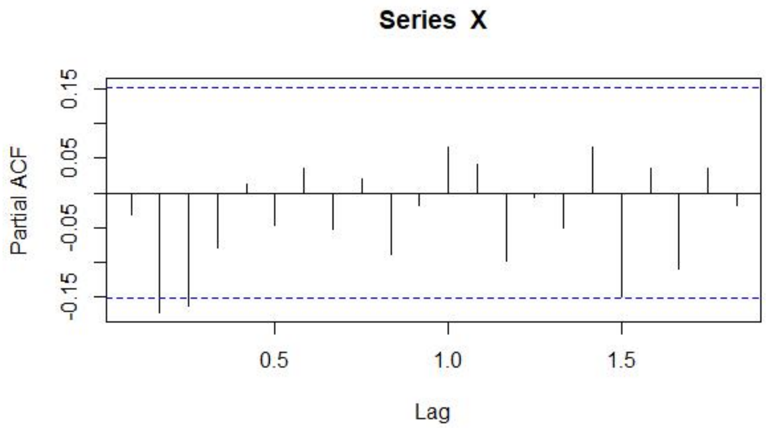
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -6.2208, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

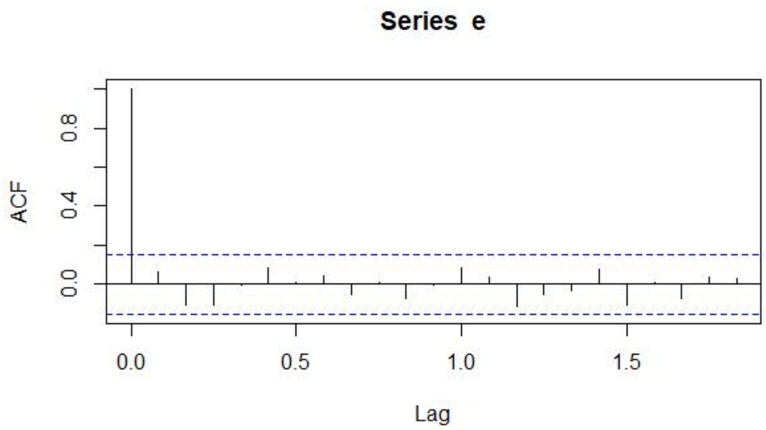
ARIMA(X,order = c (1, 1, 1)), so here we have taken p=1, d=1, q=1

Coefficients:

	ar1	ma1
	0.6533	-0.7895
s.e.	0.1673	0.1325

sigma^2 estimated as 1519: log likelihood = -848.73, aic = 1703.46

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.63662, df = 1, p-value = 0.4249

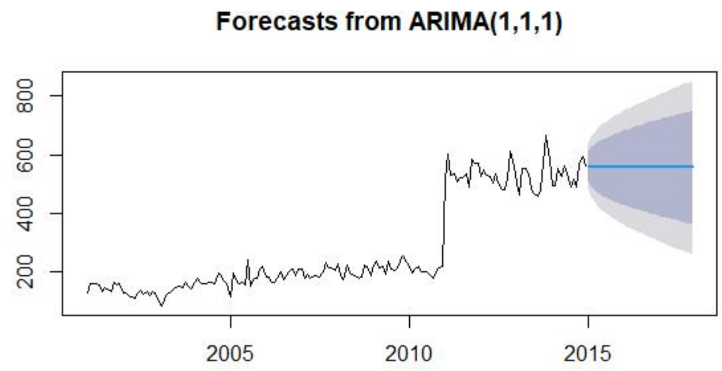
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	561.1262	559.2487	558.0221	557.2208	556.6972	556.3552	556.1317	555.9858
2016	555.7434	555.7320	555.7246	555.7198	555.7166	555.7145	555.7132	555.7123
2017	555.7108	555.7108	555.7107	555.7107	555.7107	555.7107	555.7107	555.7107

	Sep	Oct	Nov	Dec
2015	555.8904	555.8281	555.7874	555.7608
2016	555.7117	555.7114	555.7111	555.7110
2017	555.7107	555.7107	555.7107	555.7107

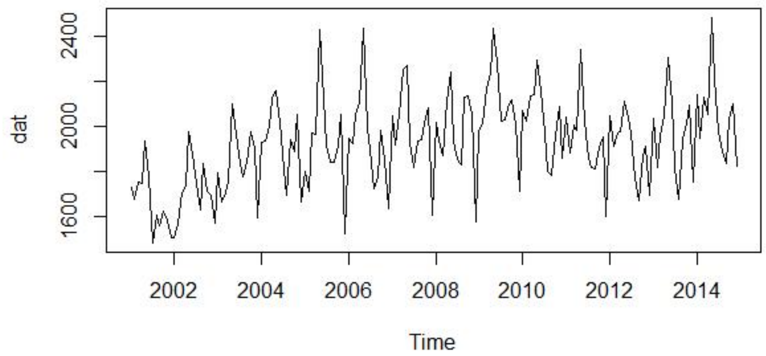
Plot for forecast is given below:



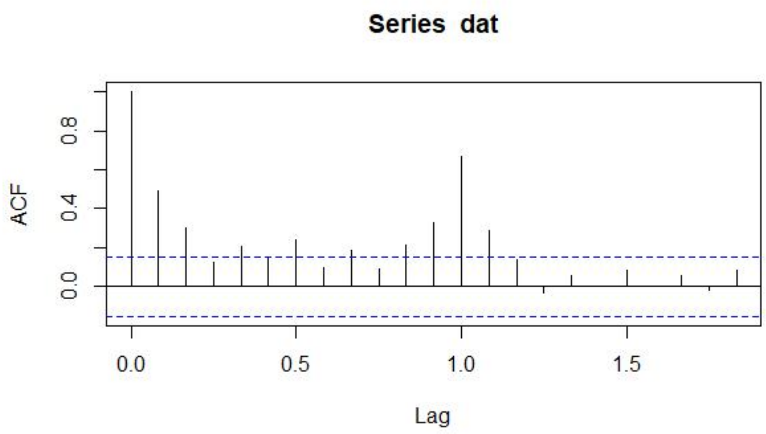
➢ **Rajasthan :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

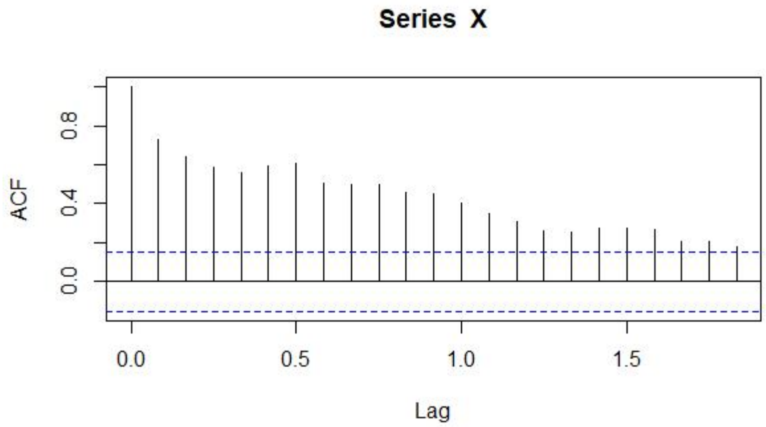


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

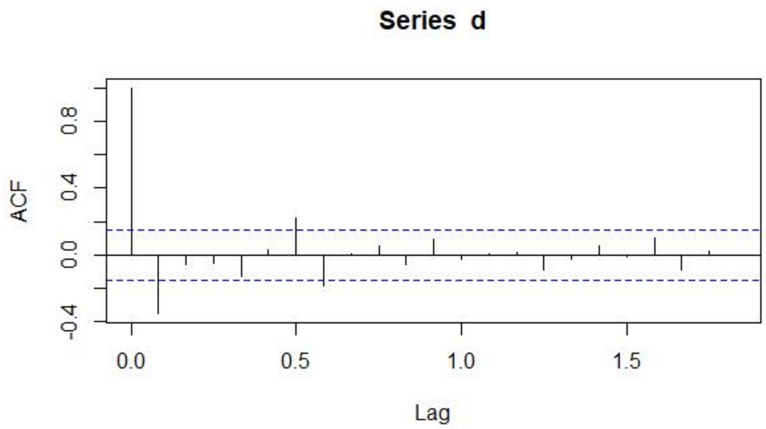


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data’s acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -2.2659, Lag order = 5, p-value = 0.4653

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

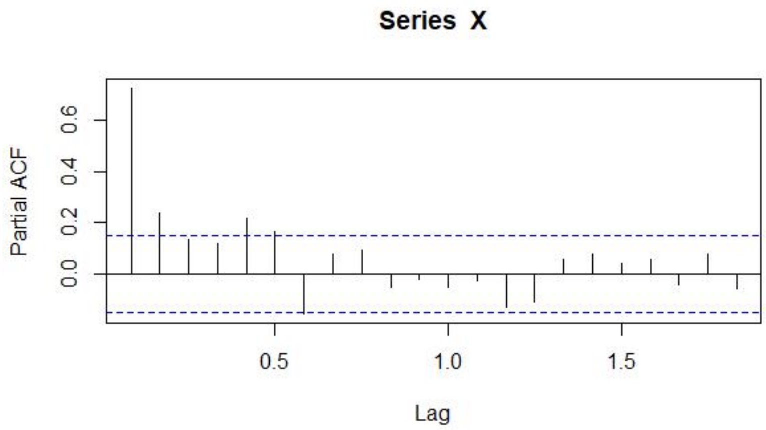
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -6.4138, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

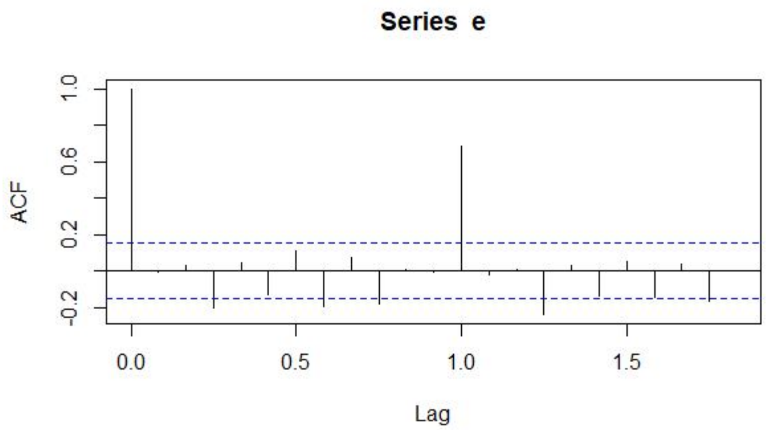
ARIMA(x = dat, order = c(1, 1, 1)), so here we have taken p=1, d=1, q=1

Coefficients:

	ar1	ma1
	0.3281	-0.9299
s.e.	0.0812	0.0282

sigma^2 estimated as 28946: log likelihood = -1095.47, aic = 2196.94

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0093727, df = 1, p-value = 0.9229

There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

Jan Feb Mar Apr May Jun Jul Aug

2015 1934.384 1970.927 1982.917 1986.851 1988.141 1988.565 1988.704 1988.749

2016 1988.771 1988.771 1988.771 1988.771 1988.771 1988.771 1988.771 1988.771

2017 1988.771 1988.771 1988.771 1988.771 1988.771 1988.771 1988.771 1988.771

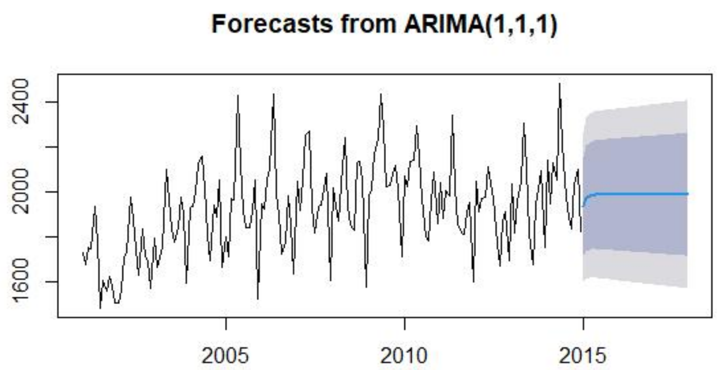
Sep Oct Nov Dec

2015 1988.764 1988.769 1988.771 1988.771

2016 1988.771 1988.771 1988.771 1988.771

2017 1988.771 1988.771 1988.771 1988.771

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

Jan Feb Mar Apr May Jun Jul Aug

2015 1953.348 1901.645 2025.227 2085.116 2322.118 2096.086 1919.254 1849.188

2016 2007.735 1919.489 2031.082 2087.037 2322.748 2096.293 1919.322 1849.210

2017 2007.736 1919.489 2031.082 2087.037 2322.748 2096.293 1919.322 1849.210

Sep Oct Nov Dec

2015 1925.587 1997.727 2011.850 1697.165

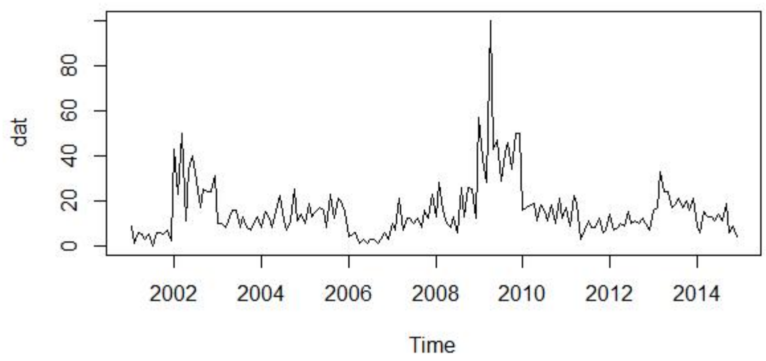
2016 1925.595 1997.729 2011.851 1697.165

2017 1925.595 1997.729 2011.851 1697.165

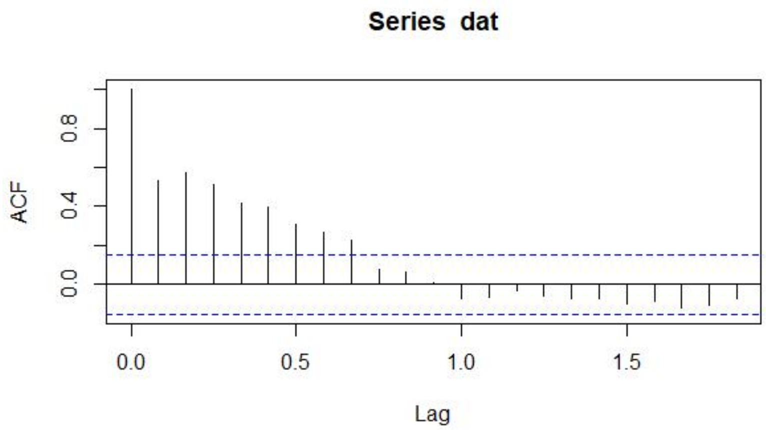
➢ **Sikkim :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

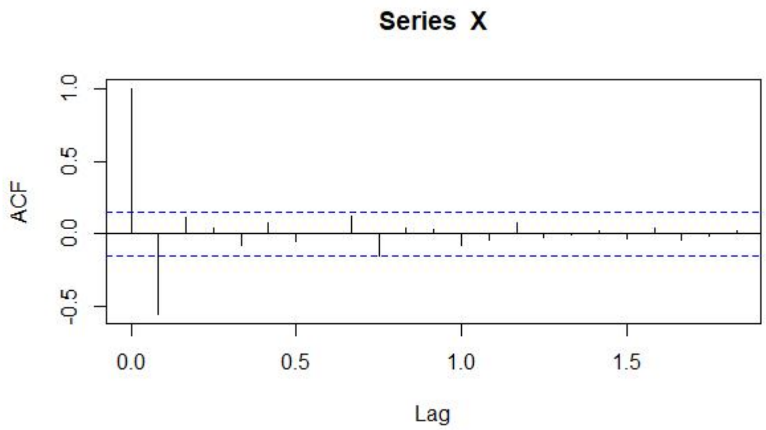
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -2.9213, Lag order = 5, p-value = 0.1917

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

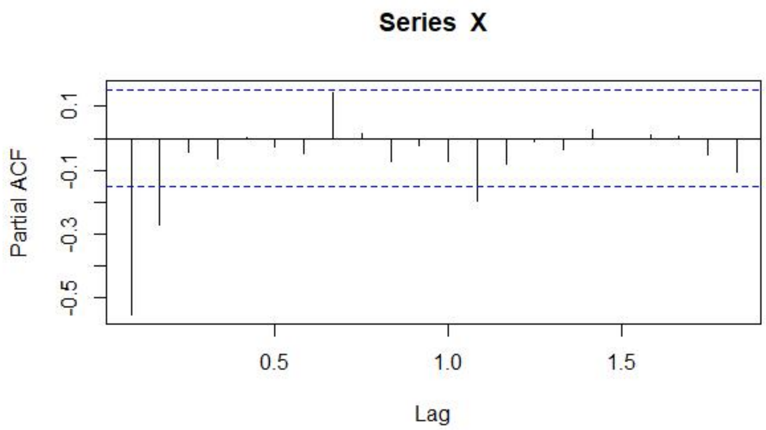
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -6.2047, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

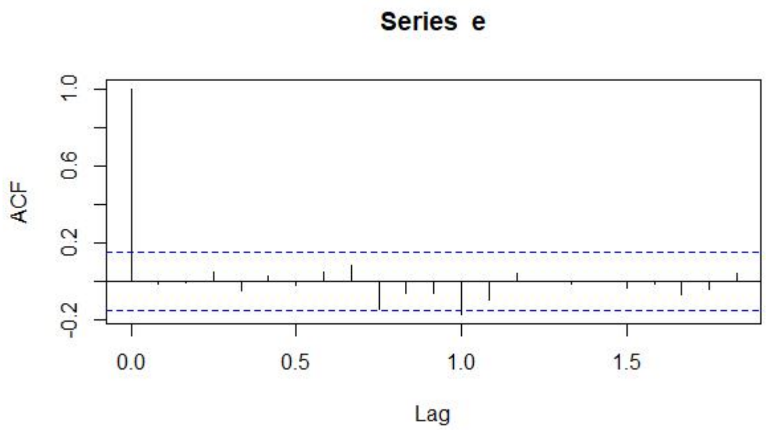
ARIMA(X,order = c (1, 1, 1)), so here we have taken p=1, d=1, q=1

Coefficients:

	ar1	ma1
	-0.271	-0.4462
s.e.	0.113	0.1035

sigma^2 estimated as 96.59: log likelihood = -618.86, aic = 1243.72

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.011997, df = 1, p-value = 0.9128

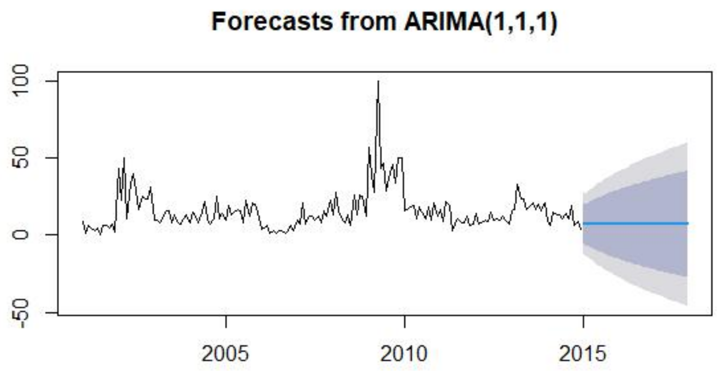
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	8.030272	6.938194	7.234113	7.153928	7.175656	7.169768	7.171364	7.170931
2016	7.171024	7.171024	7.171024	7.171024	7.171024	7.171024	7.171024	7.171024
2017	7.171024	7.171024	7.171024	7.171024	7.171024	7.171024	7.171024	7.171024

	Sep	Oct	Nov	Dec
2015	7.171049	7.171017	7.171025	7.171023
2016	7.171024	7.171024	7.171024	7.171024
2017	7.171024	7.171024	7.171024	7.171024

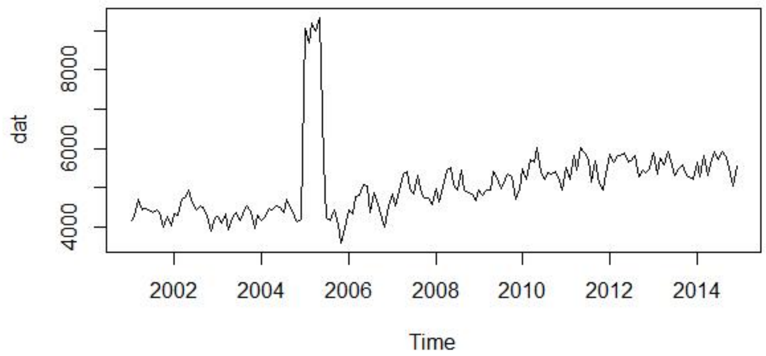
Plot for forecast is given below:



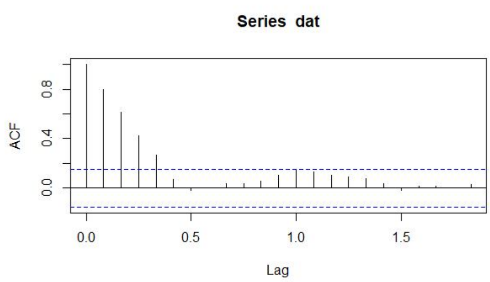
- **Tamil Nadu :**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is stationary.



So, now we can say that the data is stationary, augmented Dickey Fuller test also showed that

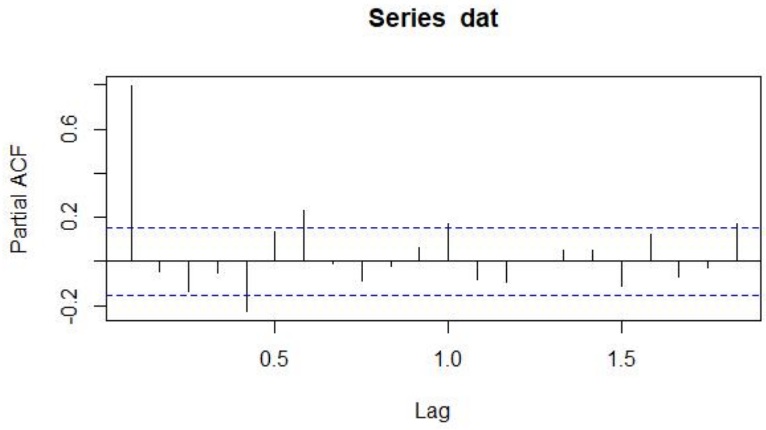
Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -5.2665, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=0.

Fitted model by checking AIC is:

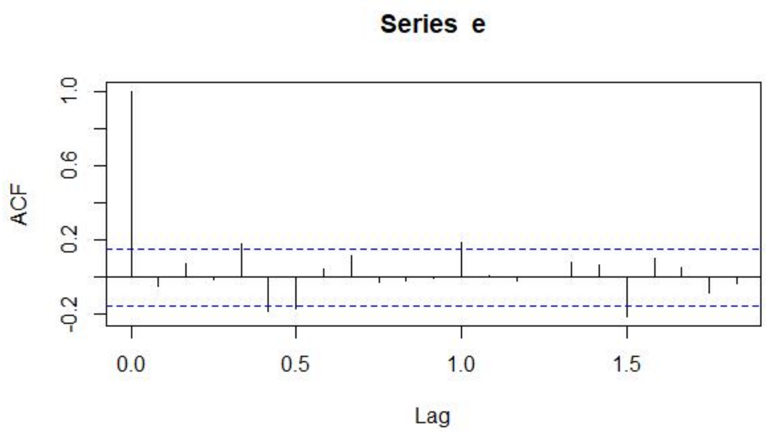
ARIMA(X,order = c (2, 0, 1)), so here we have taken p=2, d=0, q=1

Coefficients:

	ar1	ar2	ma1	intercept
	1.4856	-0.5856	-0.6182	5067.5052
s.e.	0.1902	0.1498	0.1992	157.7013

sigma^2 estimated as 291726: log likelihood = -1295.93, aic = 2601.85

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.41152, df = 1, p-value = 0.5212

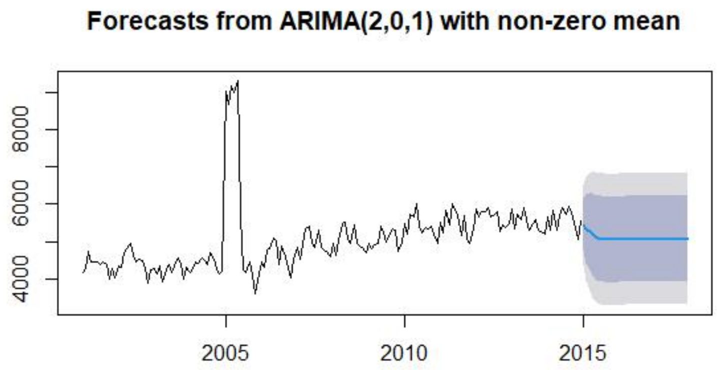
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	5448.793	5344.368	5255.529	5184.701	5131.503	5093.949	5069.312	5054.704
2016	5052.207	5055.752	5059.004	5061.758	5063.946	5065.583	5066.734	5067.485
2017	5068.087	5067.978	5067.867	5067.766	5067.681	5067.613	5067.563	5067.528

	Sep	Oct	Nov	Dec
2015	5047.429	5045.176	5046.090	5048.766
2016	5067.927	5068.143	5068.206	5068.173
2017	5067.505	5067.491	5067.485	5067.483

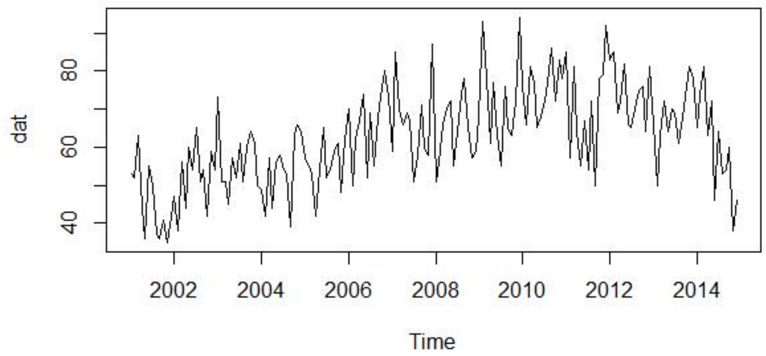
Plot for forecast is given below:



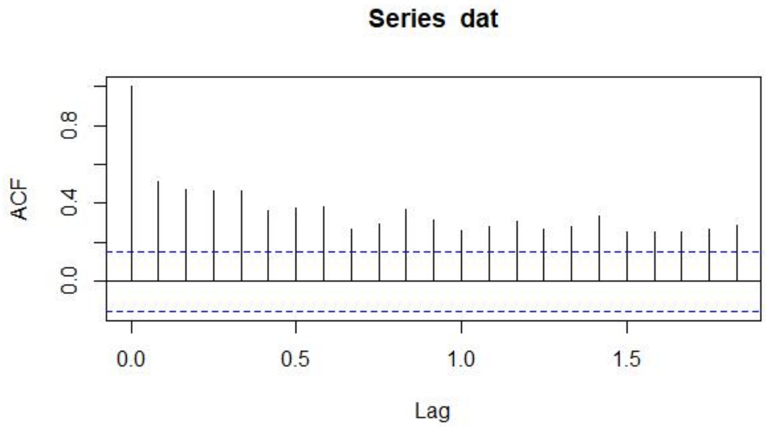
- **<u>Tripura :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

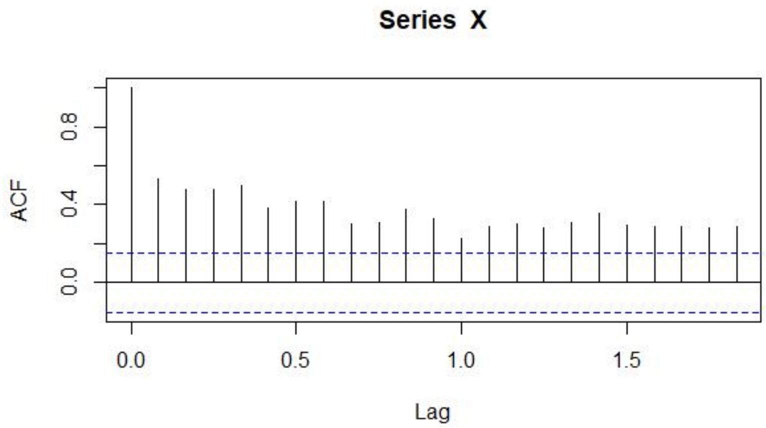


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.

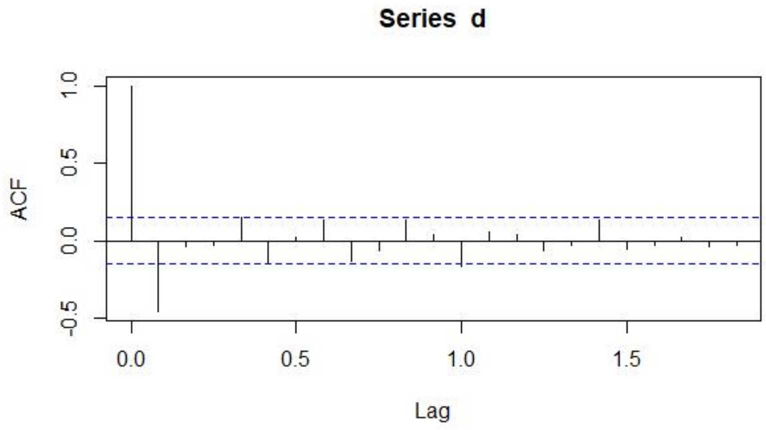


So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data’s acf plot is:



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -1.9159, Lag order = 5, p-value = 0.6114

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

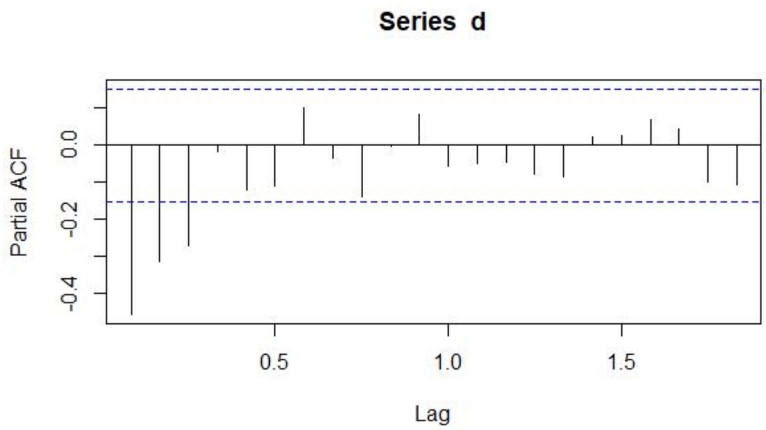
Augmented Dickey-Fuller Test

data: d

Dickey-Fuller = -7.769, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

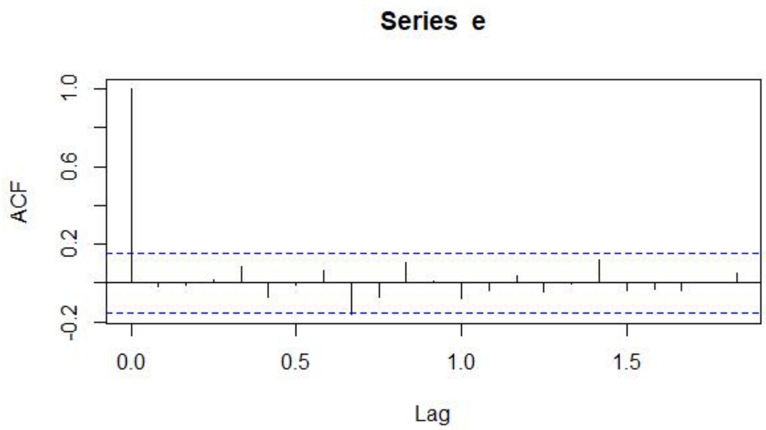
ARIMA(x = dat, order = c(1, 1, 2)), so here we have taken p=1, d=1, q=2

Coefficients:

	ar1	ma1	ma2
	0.7191	-1.4695	0.5111
s.e.	0.2408	0.2582	0.2219

sigma^2 estimated as 98.61: log likelihood = -620.87, aic = 1249.75

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.041153, df = 1, p-value = 0.8392

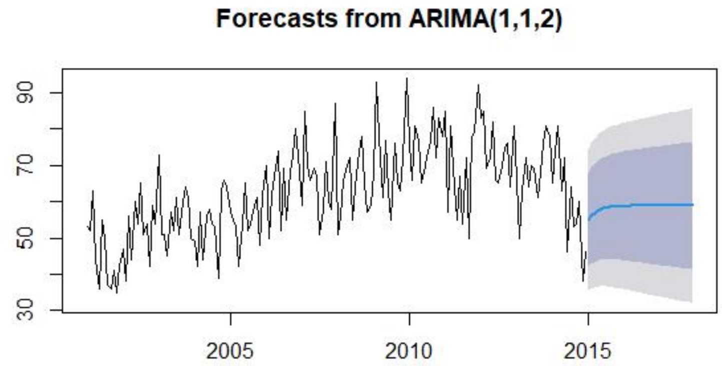
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	54.93212	56.02470	56.81035	57.37531	57.78157	58.07371	58.28378	58.43484
2016	58.74718	58.76807	58.78309	58.79389	58.80165	58.80724	58.81126	58.81414
2017	58.82011	58.82051	58.82080	58.82101	58.82116	58.82126	58.82134	58.82139

	Sep	Oct	Nov	Dec
2015	58.54347	58.62158	58.67775	58.71814
2016	58.81622	58.81771	58.81879	58.81956
2017	58.82143	58.82146	58.82148	58.82150

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

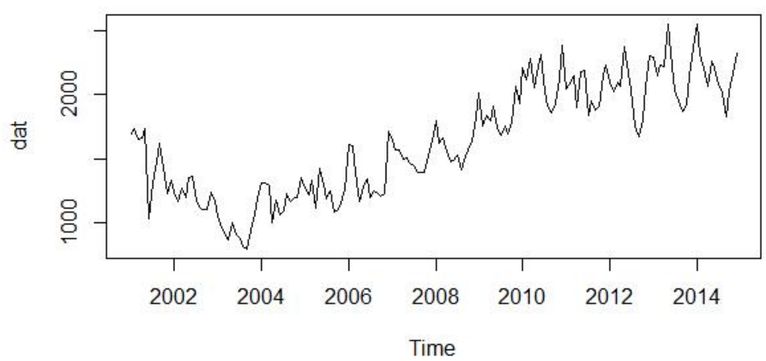
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	56.63244	54.01348	59.76708	53.98269	59.70304	54.28044	55.38794	56.04542
2016	60.44750	56.75685	61.73982	55.40126	60.72313	55.01397	55.91542	56.42472
2017	60.52043	56.80930	61.77753	55.42838	60.74263	55.02799	55.92551	56.43197

	Sep	Oct	Nov	Dec
2015	57.56430	59.44690	60.01909	65.43449
2016	57.83705	59.64303	60.16013	65.53590
2017	57.84227	59.64678	60.16283	65.53784

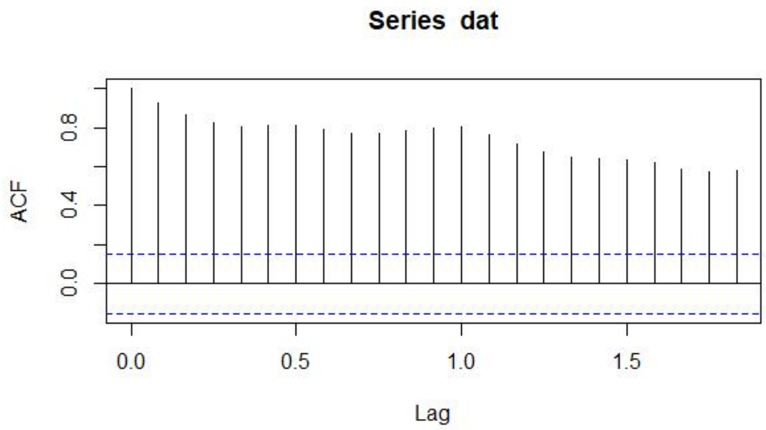
➢ **<u>Uttar Pradesh :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

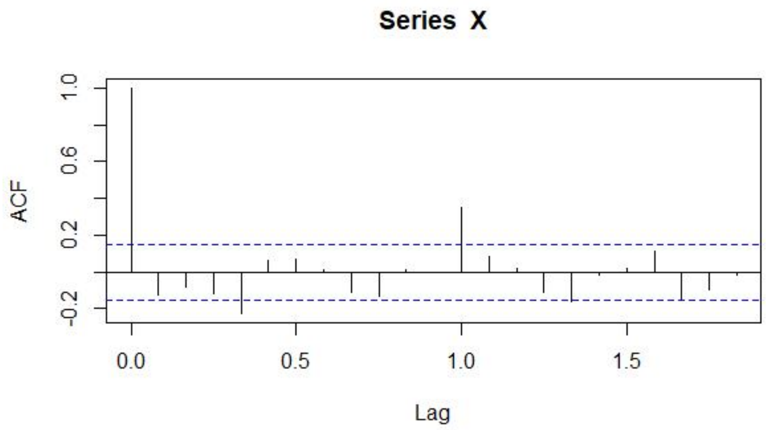
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -3.1535, Lag order = 5, p-value = 0.09793

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

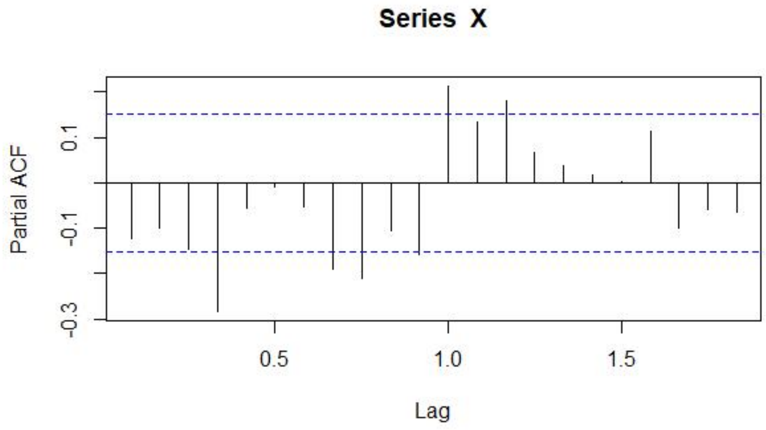
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -7.068, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

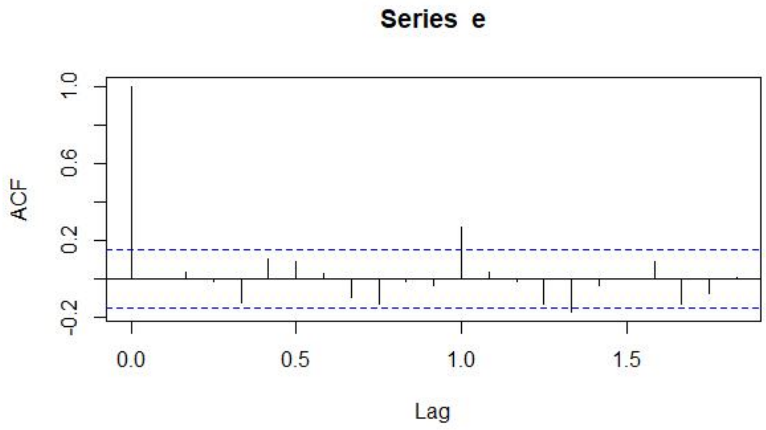
ARIMA(X,order = c (2, 1, 2)), so here we have taken p=2, d=1, q=2

Coefficients:

	ar1	ar2	ma1	ma2
	1.5977	-0.7238	-1.9369	1.0000
s.e.	0.0551	0.0566	0.0420	0.0431

sigma^2 estimated as 20575: log likelihood = -1069.24, aic = 2148.49

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.0011028, df = 1, p-value = 0.9735

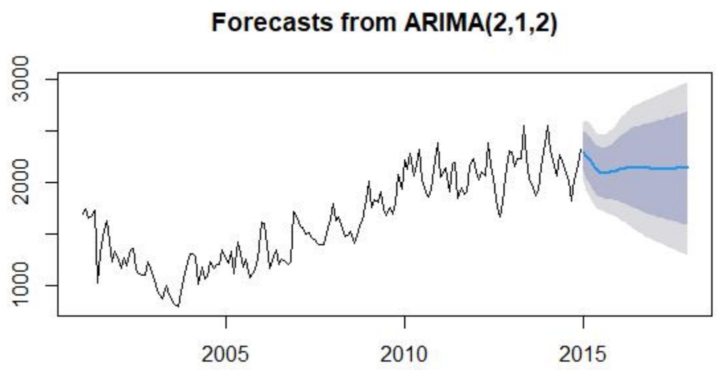
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	2299.238	2259.781	2215.385	2173.008	2137.432	2111.261	2095.195	2088.466
2016	2125.371	2133.643	2139.685	2143.351	2144.836	2144.555	2143.031	2140.800
2017	2131.867	2131.540	2131.624	2131.993	2132.523	2133.103	2133.645	2134.092

	Sep	Oct	Nov	Dec
2015	2089.343	2095.615	2105.000	2115.457
2016	2138.338	2136.020	2134.097	2132.702
2017	2134.413	2134.604	2134.675	2134.651

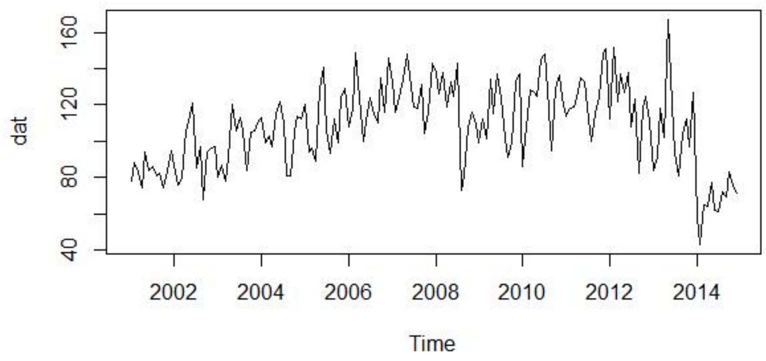
Plot for forecast is given below:



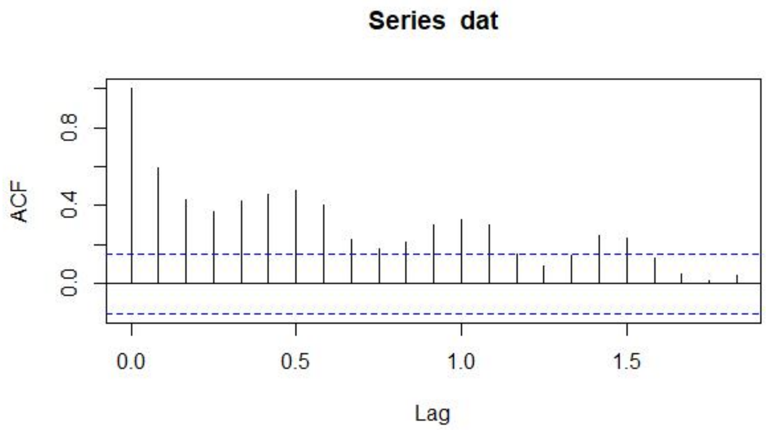
➢ **<u>Uttarakhand :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

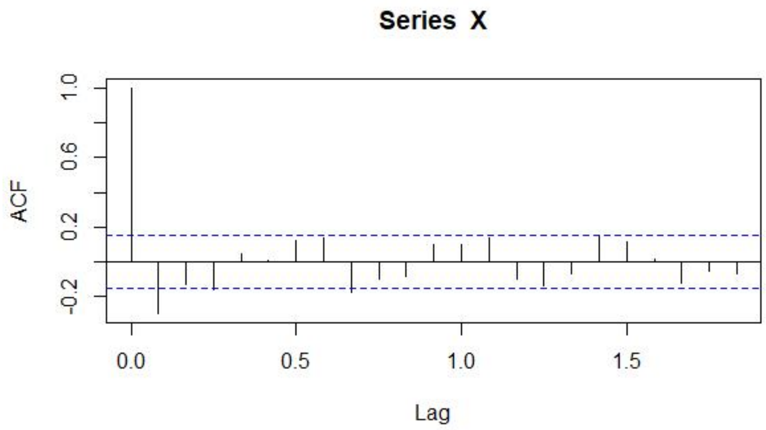
Graphical representation of whole data:



Now plotting acf plot of data we can say that the data is not stationary.



After differentiating the data we again take the acf plot.



So, now we can say that the data is not stationary, augmented Dickey Fuller test also showed that

Augmented Dickey-Fuller Test

data: dat

Dickey-Fuller = -1.2973, Lag order = 5, p-value = 0.8696

alternative hypothesis: stationary

Now, we will take the diff to make it stationary.

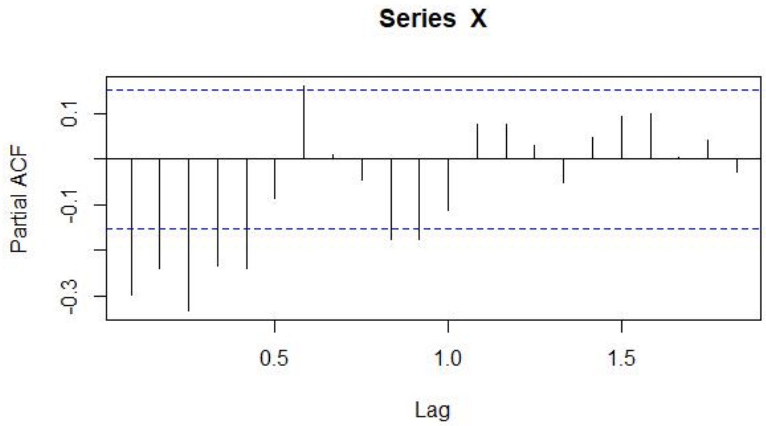
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -9.0443, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

The pacf plot for the differenced data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

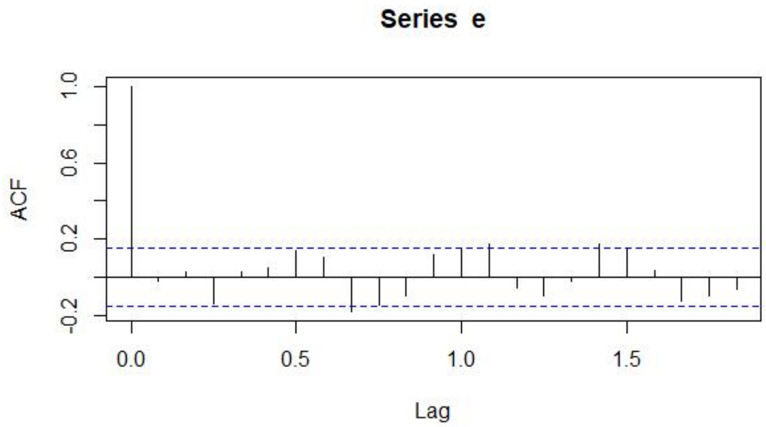
ARIMA(X,order = c (2, 1, 1)), so here we have taken p=2, d=1, q=1

Coefficients:

	ar1	ar2	ma1
	0.2327	-0.1377	-0.7971
s.e.	0.0932	0.0850	0.0593

sigma^2 estimated as 295: log likelihood = -712.26, aic = 1432.52

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.081361, df = 1, p-value = 0.7755

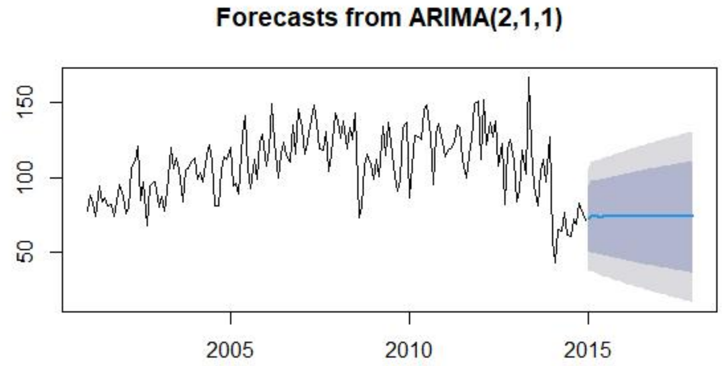
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	72.66432	73.73992	73.76103	73.61786	73.58164	73.59292	73.60054	73.60075
2016	73.59963	73.59962	73.59962	73.59962	73.59962	73.59962	73.59962	73.59962
2017	73.59962	73.59962	73.59962	73.59962	73.59962	73.59962	73.59962	73.59962

	Sep	Oct	Nov	Dec
2015	73.59976	73.59949	73.59957	73.59962
2016	73.59962	73.59962	73.59962	73.59962
2017	73.59962	73.59962	73.59962	73.59962

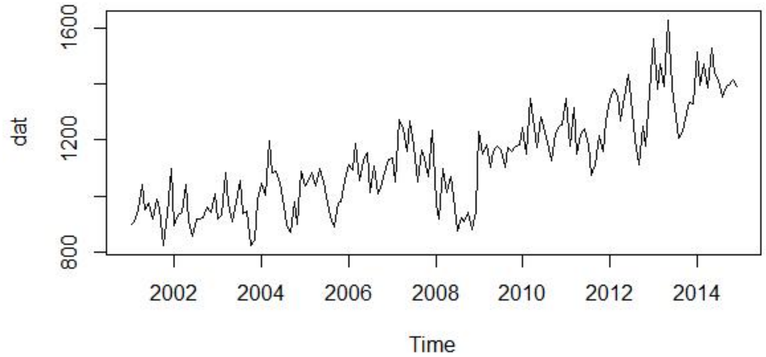
Plot for forecast is given below:



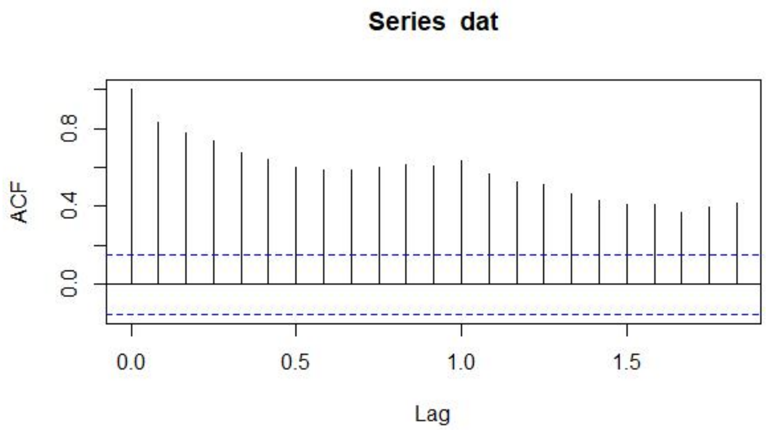
- **<u>West Bengal :</u>**

We have the monthly road accidents data for 2001 to 2014, so there are 168 data points to fit in ARIMA model.

Graphical representation of whole data:

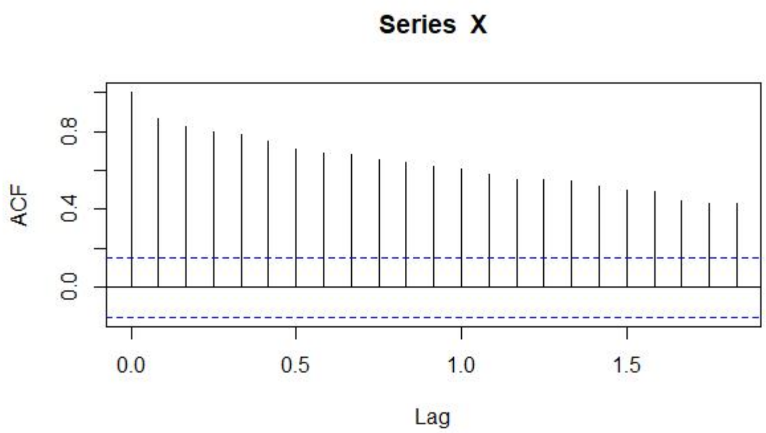


Now plotting acf plot of data we can say that the data is not stationary and may be there is seasonality present.



So, now we can say that the data has seasonality, so we can decomposed the data and removed the seasonal part. Now, we will proceed with the new data (deseasosonalized data).

After deseasosonalized the new data's acf plot is:



So, now we can say that the data is stationary, augmented Dickey Fuller test also showed that

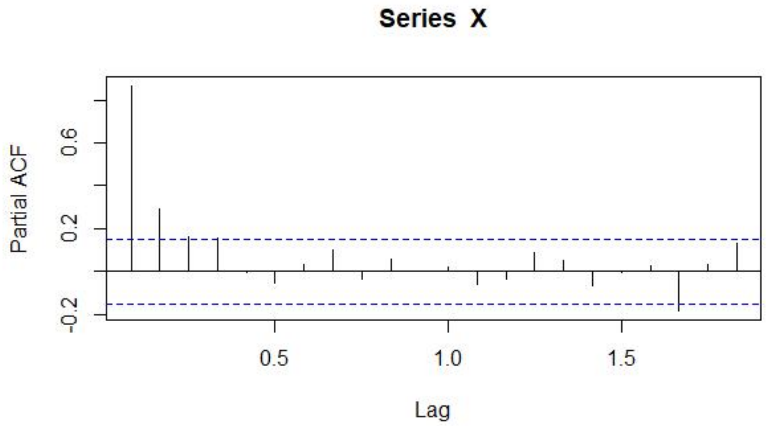
Augmented Dickey-Fuller Test

data: X

Dickey-Fuller = -3.4701, Lag order = 5, p-value = 0.04732

alternative hypothesis: stationary

The pacf plot for the new data:



Therefore the data is stationary at d=1.

Fitted model by checking AIC is:

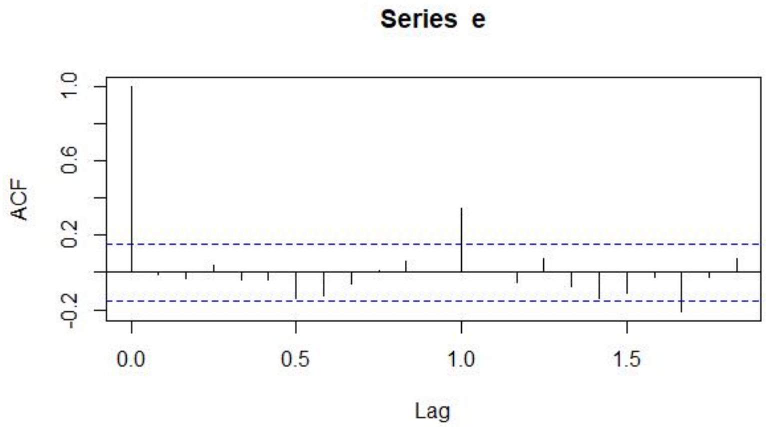
ARIMA(x = dat, order = c(2, 0, 3)), so here we have taken p=2, d=0, q=3

Coefficients:

	ar1	ar2	ma1	ma2	ma3	intercept
	1.5799	-0.5819	-1.1005	0.2754	-0.0806	1147.612
s.e.	0.1794	0.1783	0.1868	0.1532	0.0839	167.959

sigma^2 estimated as 7445: log likelihood = -988.44, aic = 1990.89

For diagnostic checking for model, we have used the residual's acf plot and Ljung Box test .



Box-Ljung test

data: e

X-squared = 0.030691, df = 1, p-value = 0.8609

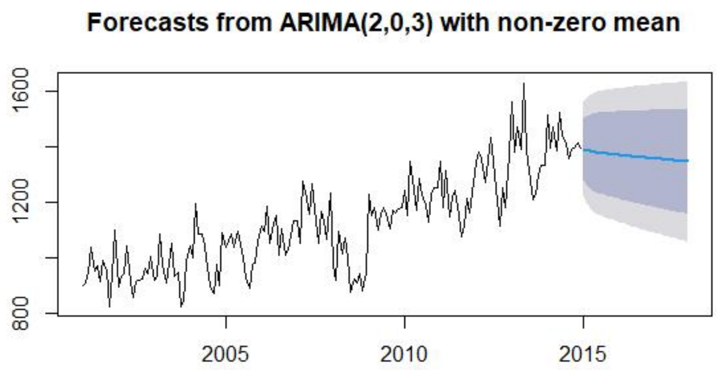
There fore, we can say that the model is good,as the acf plot shows that there is single spike at h=0 and h>0 it is insignificant. And the p-value for Ljung Box test is >0.05, so we can say that the model is good fit.

Now we will take the prediction for 3 years from the fitted model.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	1392.456	1387.575	1384.596	1382.375	1380.601	1379.091	1377.736	1376.476
2016	1370.714	1369.606	1368.505	1367.411	1366.323	1365.241	1364.165	1363.094
2017	1357.818	1356.779	1355.744	1354.715	1353.691	1352.672	1351.658	1350.649

	Sep	Oct	Nov	Dec
2015	1375.272	1374.105	1372.960	1371.831
2016	1362.028	1360.968	1359.913	1358.863
2017	1349.645	1348.646	1347.651	1346.662

Note that, this plot is for deseasonalized data:



Now, the original prediction are, i.e. after adding the seasonal part is given below,

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2015	1448.614	1384.569	1485.705	1400.863	1428.487	1417.335	1359.657	1306.996
2016	1426.872	1366.600	1469.615	1385.899	1414.209	1403.486	1346.085	1293.614
2017	1413.976	1353.773	1456.854	1373.203	1401.576	1390.916	1333.578	1281.169

	Sep	Oct	Nov	Dec
2015	1293.385	1320.349	1309.637	1399.476
2016	1280.141	1307.212	1296.590	1386.508
2017	1267.758	1294.890	1284.329	1374.307

Conclusion

The statistical analysis of road accidents across different states in India highlights significant variations in accident trends. States like Tamil Nadu, Uttar Pradesh, and Maharashtra consistently record higher numbers of accidents, while smaller states and regions with less vehicular density report comparatively lower figures.

Predictive models developed from historical accident data provide crucial insights into future predictions. These models enable the identification of high-risk states and peak periods for accidents. By accurately forecasting accident patterns.

- **Andaman and Nicobar Islands:** Analysis of road accident data in Andaman and Nicobar Islands indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Andhra Pradesh:** Analysis of road accident data in Andhra Pradesh reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Arunachal Pradesh:** Analysis of road accident data in Arunachal Pradesh reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Assam:** Analysis of road accident data in Assam indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Bihar:** Analysis of road accident data in Bihar reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be a slight decrease in accidents in the year 2015,2016,2017.
- **Chandigarh:** Analysis of road accident data in Chandigarh indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be a slight rise in accidents in the year 2015,2016,2017.
- **Chhattisgarh:** Analysis of road accident data in Chhattisgarh indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be a high rise in accidents in the year 2015,2016,2017.
- **Dadar and Nagar Haveli:** Analysis of road accident data in Dadar and Nagar Haveli indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be no change in the rate of accidents in the year 2015,2016,2017.
- **Daman and Diu:** Analysis of road accident data in Daman and Diu reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy.

Forecasts indicate that there will be a slight decrease in accidents in the year 2015,2016,2017.

- **Delhi(UT):** Analysis of road accident data in Delhi(UT) indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be no change in the rate of accidents in the year 2015,2016,2017.
- **Goa:** Analysis of road accident data in Goa reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Gujarat:** Analysis of road accident data in Gujarat reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be decrease in accidents in the year 2015,2016,2017.
- **Haryana:** Analysis of road accident data in Haryana indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Himachal Pradesh:** Analysis of road accident data in Himachal Pradesh reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be high rise in accidents in the year 2015,2016,2017.
- **Jammu and Kashmir:** Analysis of road accident data in Jammu and Kashmir reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be no change in the rate of accidents in the year 2015,2016,2017.
- **Jharkhand:** Analysis of road accident data in Jharkhand reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be high rise in accidents in the year 2015,2016,2017.
- **Karnataka:** Analysis of road accident data in Karnataka reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be decrease in accidents in the year 2015,2016,2017.
- **Kerala:** Analysis of road accident data in Kerala reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be decrease in accidents in the year 2015,2016,2017.
- **Lakshadweep:** Analysis of road accident data in Lakshadweep indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Madhya Pradesh:** Analysis of road accident data in Madhya Pradesh reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be rise in accidents in the year 2015,2016,2017.
- **Maharashtra:** Analysis of road accident data in Maharashtra reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be decrease in accidents in the year 2015,2016,2017.

- **Manipur:** Analysis of road accident data in Manipur reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be steep decrease in accidents in the year 2015,2016,2017.
- **Meghalaya:** Analysis of road accident data in Meghalaya indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be no change in the rate of accidents in the year 2015,2016,2017.
- **Mizoram:** Analysis of road accident data in Mizoram indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be no change in the rate of accidents in the year 2015,2016,2017.
- **Nagaland:** Analysis of road accident data in Nagaland indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be decrease in accidents in the year 2015,2016,2017.
- **Odisha:** Analysis of road accident data in Odisha reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be steep decrease in accidents in the year 2015,2016,2017.
- **Puducherry:** Analysis of road accident data in Puducherry indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Punjab:** Analysis of road accident data in Punjab indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Rajasthan:** Analysis of road accident data in Rajasthan reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be no change in the rate of accidents in the year 2015,2016,2017.
- **Sikkim:** Analysis of road accident data in Sikkim indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be decrease in accidents in the year 2015,2016,2017.
- **Tamil Nadu:** Analysis of road accident data in Tamil Nadu indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be no change in the rate of accidents in the year 2015,2016,2017.
- **Tripura:** Analysis of road accident data in Tripura reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be rise in accidents in the year 2015,2016,2017.
- **Uttar Pradesh:** Analysis of road accident data in Uttar Pradesh indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be a rise in accidents in the year 2015,2016,2017.
- **Uttarakhand:** Analysis of road accident data in Uttarakhand indicates the absence of seasonality. A non-seasonal ARIMA model provided the best forecast. Forecasts indicate that there will be decrease in accidents in the year 2015,2016,2017.

- **West Bengal:** Analysis of road accident data in West Bengal reveals significant seasonality. ARIMA model with seasonal adjustments provided the best forecast accuracy. Forecasts indicate that there will be steep decrease in accidents in the year 2015,2016,2017.

Appendix

Original dataset:

STATE/UT	YEAR	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUGU	SEPT	OCT	NOV	DEC	TOTAL
A & N Islands	2001	8	23	15	15	14	19	14	19	7	12	13	22	181
A & N Islands	2002	12	10	14	16	10	7	16	11	23	21	11	17	168
A & N Islands	2003	19	13	15	13	13	12	8	16	17	25	14	15	180
A & N Islands	2004	21	14	22	17	13	18	16	19	16	20	15	24	215
A & N Islands	2005	19	21	22	17	13	19	21	14	15	19	10	16	206
A & N Islands	2006	21	13	4	22	9	14	12	14	8	14	6	18	155
A & N Islands	2007	17	16	12	22	12	14	8	10	11	7	11	12	152
A & N Islands	2008	17	22	15	16	15	17	13	11	13	17	11	24	191
A & N Islands	2009	16	23	23	21	21	19	24	25	31	22	20	26	271
A & N Islands	2010	16	30	28	15	29	24	22	18	25	30	27	21	285
A & N Islands	2011	24	10	19	24	13	28	17	18	25	17	18	22	235
A & N Islands	2012	25	15	24	24	18	16	17	18	18	25	17	19	236
A & N Islands	2013	24	23	16	15	13	16	14	25	14	15	14	11	200
A & N Islands	2014	25	13	19	19	18	15	15	16	15	23	18	22	218
Andhra Pradesh	2001	2204	2437	2405	2351	2550	2284	2025	2077	2070	2276	2122	2387	27188
Andhra Pradesh	2002	2492	2453	2835	2786	3195	2880	2645	2607	2555	2624	2646	2859	32577
Andhra Pradesh	2003	2783	2569	2870	2635	3265	2924	2657	2934	2767	2881	3037	3215	34537
Andhra Pradesh	2004	3019	3131	3211	3100	3257	2942	2827	3079	2972	3041	3129	3370	37078
Andhra Pradesh	2005	3189	3193	3182	3056	3612	3247	2907	3028	2742	2928	2975	3230	37289
Andhra Pradesh	2006	3568	3224	3496	3634	3962	3400	3334	3311	3232	3306	3268	3588	41323
Andhra Pradesh	2007	3978	3530	3728	3842	4099	3594	3519	3348	3246	3447	3617	3646	43594
Andhra Pradesh	2008	3594	3468	3848	3967	3811	3391	3260	3324	3169	3352	3319	3603	42106
Andhra Pradesh	2009	3682	3494	3775	3450	4048	3763	3412	3488	3017	3253	3298	3331	42011
Andhra Pradesh	2010	3515	3434	3749	3857	3960	3765	3206	3416	3115	3439	3397	3575	42428
Andhra Pradesh	2011	3540	3195	3584	3396	3916	3793	3237	3106	3067	3398	3442	3392	41066
Andhra Pradesh	2012	3347	3390	3693	3589	3250	3187	3160	3177	2893	3205	2985	3468	39344
Andhra Pradesh	2013	3732	3482	3715	3648	4531	3736	3018	3465	3450	3137	3432	3702	43048
Andhra Pradesh	2014	3809	3657	3641	3582	3986	3664	3167	3587	3225	3410	3346	4158	43232
Arunachal Pradesh	2001	19	17	21	16	19	14	21	14	23	17	30	25	236
Arunachal Pradesh	2002	25	16	21	23	16	19	11	17	18	21	25	23	235
Arunachal Pradesh	2003	16	24	22	13	18	17	19	21	23	18	16	22	229
Arunachal Pradesh	2004	23	24	23	11	15	11	14	8	21	29	17	21	217
Arunachal Pradesh	2005	26	14	29	12	17	20	20	14	19	21	24	21	237
Arunachal Pradesh	2006	14	20	17	19	20	31	11	25	18	25	21	22	243
Arunachal Pradesh	2007	22	20	26	19	25	19	14	15	22	18	11	19	230
Arunachal Pradesh	2008	20	21	18	16	33	24	13	18	27	20	24	27	261
Arunachal Pradesh	2009	20	21	18	16	33	24	13	18	27	20	24	27	261
Arunachal Pradesh	2010	31	33	20	26	28	23	14	15	18	20	23	29	280
Arunachal Pradesh	2011	22	24	30	22	29	9	28	22	16	25	15	21	263
Arunachal Pradesh	2012	17	21	24	16	11	13	20	18	14	16	20	14	204
Arunachal Pradesh	2013	23	33	16	23	25	27	27	28	24	25	29	28	308
Arunachal Pradesh	2014	28	10	10	21	9	12	13	14	14	15	22	17	185
Assam	2001	141	127	202	173	176	137	140	160	132	140	132	146	1806
Assam	2002	196	210	185	159	182	163	144	237	180	180	222	185	2243
Assam	2003	194	143	193	180	169	180	162	167	205	188	180	184	2145
Assam	2004	183	172	170	160	176	168	118	170	157	167	182	179	2002
Assam	2005	307	334	317	287	306	308	297	354	286	311	287	262	3656
Assam	2006	297	333	365	318	363	352	370	289	335	325	378	355	4080
Assam	2007	383	358	392	395	375	331	299	305	296	308	275	278	3995
Assam	2008	375	381	399	363	339	326	371	370	344	308	346	340	4262
Assam	2009	388	385	400	377	394	376	393	331	387	405	367	382	4585
Assam	2010	418	450	496	452	396	416	455	454	504	525	451	468	5485
Assam	2011	554	545	607	515	518	557	502	515	532	551	590	583	6569
Assam	2012	590	599	605	601	567	536	486	458	418	544	543	588	6535
Assam	2013	634	555	603	632	595	580	535	552	601	697	635	592	7211
Assam	2014	658	615	689	612	566	531	573	528	460	670	638	604	7144
Bihar	2001	325	361	474	336	528	417	329	336	317	289	317	305	4334
Bihar	2002	380	366	401	391	560	477	362	303	227	290	326	289	4372
Bihar	2003	337	340	368	359	440	429	326	241	255	263	258	286	3902
Bihar	2004	296	349	407	308	413	392	253	258	320	290	297	307	3890
Bihar	2005	262	254	354	340	463	381	310	263	241	255	301	322	3746
Bihar	2006	306	348	395	389	439	386	363	341	320	381	380	334	4382
Bihar	2007	443	464	505	518	635	551	438	375	347	427	484	444	5631
Bihar	2008	447	485	612	583	621	529	441	465	422	487	591	497	6180
Bihar	2009	543	706	754	595	925	911	712	592	574	634	699	721	8366
Bihar	2010	610	698	722	650	833	1034	723	689	546	606	665	665	8441
Bihar	2011	750	808	964	737	964	880	769	644	618	759	773	689	9355
Bihar	2012	799	905	915	1027	909	1003	822	787	728	763	837	825	10320
Bihar	2013	800	792	815	813	1269	913	882	758	735	765	845	811	10198
Bihar	2014	777	742	852	714	985	1002	757	662	685	720	829	806	9531

Chandigarh	2001	48	25	34	47	48	46	34	31	53	37	38	52	493
Chandigarh	2002	33	42	47	42	39	34	45	46	39	38	41	48	494
Chandigarh	2003	43	35	38	40	36	32	29	36	39	43	41	31	443
Chandigarh	2004	27	36	40	23	36	26	29	34	40	37	45	38	411
Chandigarh	2005	46	38	32	27	43	44	41	50	56	52	55	44	528
Chandigarh	2006	52	42	43	42	51	42	31	43	45	46	34	50	521
Chandigarh	2007	36	39	50	39	48	35	45	48	55	60	48	33	536
Chandigarh	2008	38	34	46	45	42	39	38	46	36	42	34	37	477
Chandigarh	2009	35	32	41	25	26	29	37	35	50	38	40	36	424
Chandigarh	2010	31	39	40	35	31	33	31	30	36	54	44	52	456
Chandigarh	2011	27	46	40	37	49	35	35	27	26	45	24	50	441
Chandigarh	2012	34	41	29	40	26	30	36	43	30	35	35	33	412
Chandigarh	2013	36	27	38	43	40	34	38	31	30	31	37	24	409
Chandigarh	2014	35	22	31	40	32	25	22	32	38	31	28	30	366
Chhattisgarh	2001	428	385	403	448	504	425	367	371	400	425	393	438	4987
Chhattisgarh	2002	477	439	522	511	584	468	434	444	448	515	453	506	5801
Chhattisgarh	2003	649	565	589	647	629	613	586	524	501	595	543	582	7023
Chhattisgarh	2004	553	532	534	586	592	501	409	437	439	532	483	477	6075
Chhattisgarh	2005	472	513	537	513	566	530	558	421	435	496	483	472	5996
Chhattisgarh	2006	624	611	607	631	710	598	534	534	507	575	607	573	7111
Chhattisgarh	2007	787	632	638	729	701	630	619	544	642	660	599	656	7837
Chhattisgarh	2008	773	712	746	859	848	704	659	589	706	783	655	735	8769
Chhattisgarh	2009	721	689	643	689	710	670	639	595	617	642	659	639	7913
Chhattisgarh	2010	670	627	671	652	705	612	655	606	605	684	655	660	7802
Chhattisgarh	2011	868	795	804	725	903	793	747	660	636	770	723	771	9195
Chhattisgarh	2012	1253	1161	1175	1188	1182	1121	1103	990	1016	1107	1084	1131	13511
Chhattisgarh	2013	1384	1269	1288	1139	1209	1131	1025	1029	1005	1042	1046	1090	13657
Chhattisgarh	2014	1167	1110	1000	1020	1149	1051	934	877	910	1005	973	944	12140
D & N Haveli	2001	6	11	6	8	9	7	4	7	9	8	5	7	87
D & N Haveli	2002	10	6	3	5	12	9	6	5	6	3	9	6	80
D & N Haveli	2003	6	10	6	4	8	5	1	6	4	8	8	7	73
D & N Haveli	2004	6	10	12	11	10	8	7	5	11	14	5	12	111
D & N Haveli	2005	13	13	9	3	10	16	12	11	5	10	10	15	127
D & N Haveli	2006	9	5	13	8	5	8	9	6	10	7	9	14	103
D & N Haveli	2007	12	7	14	6	7	10	8	14	14	13	0	11	116
D & N Haveli	2008	10	15	12	8	14	8	8	9	5	7	9	11	116
D & N Haveli	2009	6	5	2	3	4	4	3	3	0	4	6	5	45
D & N Haveli	2010	9	10	4	7	5	7	10	4	2	6	7	8	79
D & N Haveli	2011	12	7	6	8	9	10	6	7	8	10	7	13	103
D & N Haveli	2012	7	11	15	7	5	7	4	5	9	7	2	6	85
D & N Haveli	2013	11	5	11	8	5	7	6	8	8	9	5	8	91
D & N Haveli	2014	6	7	7	10	9	4	8	7	7	8	8	6	87
Daman & Diu	2001	2	1	4	2	3	5	3	3	4	3	4	1	35
Daman & Diu	2002	5	3	8	6	4	4	10	1	5	5	5	12	68
Daman & Diu	2003	6	8	5	4	3	5	16	4	4	7	4	6	72
Daman & Diu	2004	8	2	2	2	3	6	5	3	2	3	8	3	47
Daman & Diu	2005	5	6	4	6	5	4	0	5	3	4	7	8	57
Daman & Diu	2006	6	7	4	9	9	2	4	3	2	5	5	1	57
Daman & Diu	2007	8	4	4	7	5	4	5	7	3	4	6	3	60
Daman & Diu	2008	3	2	9	9	3	1	2	2	3	5	7	4	50
Daman & Diu	2009	2	2	2	2	1	2	2	1	1	1	2	3	21
Daman & Diu	2010	6	2	0	2	4	4	2	2	4	1	3	6	36
Daman & Diu	2011	2	2	2	1	1	1	1	2	2	2	2	3	21
Daman & Diu	2012	7	3	5	5	3	4	5	1	2	1	7	7	50
Daman & Diu	2013	6	2	2	4	1	4	1	1	5	2	2	0	30
Daman & Diu	2014	6	4	3	3	2	2	0	5	1	5	2	6	39
Delhi (UT)	2001	707	681	786	762	810	779	763	811	780	821	798	784	9282
Delhi (UT)	2002	695	710	847	698	702	665	768	801	706	724	669	725	8710
Delhi (UT)	2003	638	645	689	714	719	747	785	803	791	859	778	701	8869
Delhi (UT)	2004	730	746	755	719	747	766	727	793	761	778	843	745	9110
Delhi (UT)	2005	777	722	850	698	815	773	806	900	831	780	835	793	9580
Delhi (UT)	2006	718	726	868	779	796	754	846	863	804	832	865	848	9699
Delhi (UT)	2007	818	809	839	786	752	757	834	768	833	1196	921	1215	10528
Delhi (UT)	2008	698	696	780	747	713	691	731	721	707	704	671	707	8566
Delhi (UT)	2009	607	632	644	646	657	565	610	707	609	639	642	656	7614
Delhi (UT)	2010	559	574	616	569	639	545	640	600	574	614	652	638	7220
Delhi (UT)	2011	562	619	631	575	606	570	633	637	589	685	593	580	7280
Delhi (UT)	2012	540	544	570	556	554	556	571	554	658	603	625	606	6937
Delhi (UT)	2013	590	553	613	589	635	612	644	593	622	687	703	728	7569
Delhi UT	2014	643	698	801	632	661	658	624	744	742	687	720	700	8310

Goa	2001	124	115	111	108	131	95	95	101	108	117	122	148	1375
Goa	2002	99	100	95	113	105	74	65	95	102	113	121	119	1201
Goa	2003	173	151	151	121	166	113	99	121	116	146	166	155	1678
Goa	2004	148	127	122	130	143	104	104	120	121	120	139	164	1542
Goa	2005	93	75	88	100	100	80	64	81	79	84	110	115	1069
Goa	2006	341	323	314	300	408	268	263	272	291	303	307	359	3749
Goa	2007	365	362	332	375	392	299	245	274	335	320	353	370	4022
Goa	2008	396	350	384	354	400	309	276	326	305	340	339	400	4179
Goa	2009	389	351	370	327	370	306	279	302	324	323	371	458	4170
Goa	2010	472	411	392	393	422	307	294	346	334	381	388	434	4574
Goa	2011	486	400	370	369	421	353	307	331	312	373	395	444	4561
Goa	2012	428	317	343	332	393	328	309	357	346	339	391	405	4288
Goa	2013	418	372	380	336	361	300	291	327	376	346	376	410	4293
Goa	2014	433	359	349	349	367	354	322	324	285	312	354	422	4230
Gujarat	2001	1657	1600	1791	1744	1935	1566	1420	1533	1448	1602	1769	1743	19808
Gujarat	2002	1652	1529	1186	1446	1840	1723	1498	1515	1434	1616	1644	1657	18740
Gujarat	2003	1616	1457	1456	1395	1780	1499	1276	1439	1404	1620	1501	1659	18102
Gujarat	2004	1669	1580	1601	1606	1771	1493	1357	1119	1471	1476	1673	1662	18478
Gujarat	2005	1586	1575	1625	1583	1928	1559	1321	1342	1292	1477	1661	1592	18541
Gujarat	2006	1706	1557	1721	1703	1950	1512	1323	1172	1423	1575	1578	1724	18944
Gujarat	2007	1898	1743	1918	1959	2169	1806	1509	1505	1613	1732	1867	1714	21433
Gujarat	2008	1843	1878	1943	1969	2091	1782	1562	1461	1471	1722	1605	1700	21027
Gujarat	2009	1742	1603	1743	1628	1937	1640	1339	1594	1613	1752	1674	1836	20101
Gujarat	2010	1892	1716	1803	1717	2042	1730	1469	1310	1391	1639	1860	1929	20498
Gujarat	2011	2626	2601	2629	2447	3056	2558	2346	2212	2169	2617	2524	2414	30199
Gujarat	2012	2549	2483	2370	2400	2486	2313	2098	2076	1948	2153	2294	2097	27267
Gujarat	2013	2357	2189	2266	2189	2405	2142	1788	1821	1848	1931	2051	2048	25035
Gujarat	2014	2023	1978	2039	1846	2127	1988	1673	1633	1506	1817	1678	1844	22152
Haryana	2001	481	506	517	543	543	520	540	509	540	542	571	561	6373
Haryana	2002	513	524	595	593	583	641	586	596	559	657	615	625	7087
Haryana	2003	468	512	534	528	581	480	545	493	542	568	662	511	6424
Haryana	2004	591	657	658	629	633	673	669	659	655	685	692	707	7908
Haryana	2005	626	514	604	598	692	634	664	682	565	704	703	696	7682
Haryana	2006	792	736	854	855	876	858	823	861	847	939	878	943	10262
Haryana	2007	825	780	839	839	881	871	859	909	873	904	834	902	10316
Haryana	2008	911	869	1001	998	951	847	899	909	910	1052	939	955	11241
Haryana	2009	808	963	982	928	973	977	931	929	875	966	1080	1015	11427
Haryana	2010	883	897	1035	936	932	933	892	925	816	998	939	975	11161
Haryana	2011	830	896	891	818	899	869	842	829	874	978	911	916	10553
Haryana	2012	759	819	865	841	841	791	799	791	761	862	870	972	9971
Haryana	2013	819	711	891	861	920	799	888	784	847	959	1025	969	10473
Haryana	2014	892	881	963	911	939	906	955	916	871	891	962	956	11043
Himachal Pradesh	2001	137	131	164	156	184	207	201	167	195	227	176	165	2110
Himachal Pradesh	2002	135	138	183	176	230	241	216	220	189	176	184	193	2281
Himachal Pradesh	2003	149	145	159	160	205	207	259	239	234	189	217	196	2359
Himachal Pradesh	2004	173	183	191	207	225	231	226	208	239	230	214	188	2515
Himachal Pradesh	2005	162	143	185	166	213	241	203	261	214	218	215	180	2401
Himachal Pradesh	2006	161	161	215	229	240	209	203	193	199	224	183	199	2416
Himachal Pradesh	2007	181	167	183	207	222	227	246	220	256	256	222	195	2582
Himachal Pradesh	2008	166	132	135	179	151	174	178	166	157	210	169	165	1982
Himachal Pradesh	2009	213	252	233	236	283	299	285	257	250	272	263	235	3078
Himachal Pradesh	2010	216	205	239	253	278	310	275	257	256	246	297	245	3077
Himachal Pradesh	2011	222	215	261	245	269	283	267	249	253	272	281	282	3099
Himachal Pradesh	2012	214	191	256	243	254	274	236	270	233	219	292	217	2899
Himachal Pradesh	2013	209	202	216	254	266	234	263	235	276	273	276	277	2981
Himachal Pradesh	2014	246	191	260	234	289	295	292	329	272	302	313	302	3325
Jammu & Kashmir	2001	362	485	392	372	483	448	429	410	405	423	394	408	5011
Jammu & Kashmir	2002	336	363	427	462	452	500	544	496	425	495	453	420	5373
Jammu & Kashmir	2003	401	380	423	468	535	550	635	551	509	498	517	421	5888
Jammu & Kashmir	2004	439	463	509	503	489	588	657	559	512	531	526	505	6281
Jammu & Kashmir	2005	444	291	384	412	484	573	559	499	525	518	508	472	5669
Jammu & Kashmir	2006	332	387	467	474	516	520	541	501	448	495	455	473	5609
Jammu & Kashmir	2007	447	409	466	485	505	573	537	527	505	508	461	451	5874
Jammu & Kashmir	2008	404	376	459	486	475	537	502	306	415	535	498	433	5426
Jammu & Kashmir	2009	367	364	514	460	558	503	557	560	563	600	515	451	6012
Jammu & Kashmir	2010	452	408	553	535	539	539	552	498	470	527	576	493	6142
Jammu & Kashmir	2011	394	430	500	448	645	657	648	612	610	622	565	534	6665
Jammu & Kashmir	2012	413	460	531	533	601	631	706	590	555	557	597	463	6637
Jammu & Kashmir	2013	418	355	464	530	646	617	596	574	542	652	582	479	6455
Jammu & Kashmir	2014	414	415	424	433	506	594	563	515	415	544	517	438	5778

Jharkhand	2001	83	76	88	95	94	70	79	88	78	81	86	77	995
Jharkhand	2002	85	41	70	91	80	69	67	80	79	75	68	75	880
Jharkhand	2003	80	42	72	90	85	70	68	89	69	70	69	76	880
Jharkhand	2004	116	99	105	122	128	102	73	125	104	106	106	109	1295
Jharkhand	2005	228	179	295	253	325	236	220	201	165	193	232	212	2739
Jharkhand	2006	322	379	400	392	335	395	400	324	312	324	359	359	4301
Jharkhand	2007	429	368	398	393	513	412	310	320	311	363	312	375	4504
Jharkhand	2008	230	203	225	278	355	333	265	201	276	299	297	297	3259
Jharkhand	2009	327	342	457	360	478	427	372	340	327	354	374	378	4536
Jharkhand	2010	252	360	357	351	383	436	365	322	333	353	336	380	4228
Jharkhand	2011	323	326	338	364	408	356	316	279	287	297	316	291	3901
Jharkhand	2012	369	374	387	457	364	396	346	355	373	381	391	432	4625
Jharkhand	2013	449	423	440	427	550	429	440	391	352	406	428	426	5161
Jharkhand	2014	399	434	418	425	454	414	391	384	341	419	373	453	4905
Karnataka	2001	2706	2600	2704	2816	3215	2783	2656	2539	2437	2582	2744	3078	32860
Karnataka	2002	2885	2860	2906	2995	3415	2987	2812	2875	2815	2908	2954	3249	35661
Karnataka	2003	3063	2986	2947	2940	3456	3199	2999	2994	3001	3147	3304	3475	37511
Karnataka	2004	3389	3357	3358	3180	3329	3066	3040	2893	3093	3011	3309	3726	38751
Karnataka	2005	3457	3196	3566	3339	4075	3354	3151	3018	3184	3237	3177	3519	40273
Karnataka	2006	3618	3278	3766	3693	4132	3356	3361	3473	3519	3621	3508	3955	43280
Karnataka	2007	3801	3613	3932	4274	4440	3878	3629	3619	3499	3830	3709	4110	46334
Karnataka	2008	4092	3922	4203	4295	4236	3792	3634	3555	3424	3728	3540	3831	46252
Karnataka	2009	3743	3525	3898	3843	4212	3906	3501	3551	3527	3527	3787	4170	45190
Karnataka	2010	4013	3972	4212	4105	4538	3942	3648	3382	3431	3512	3733	3732	46220
Karnataka	2011	3936	3676	3775	3707	4309	3781	3577	3437	3451	3566	3603	3878	44696
Karnataka	2012	3906	3749	3715	3963	3871	3687	3375	3561	3460	3519	3613	4029	44448
Karnataka	2013	3507	3322	3437	3403	3732	3247	2964	2981	3013	3185	3267	3533	39591
Karnataka	2014	3915	3649	3805	3577	4053	3772	3421	3428	3254	3517	3442	3861	43694
Kerala	2001	3199	2935	2972	2912	2887	2829	2845	3012	3060	3048	3247	3493	36439
Kerala	2002	3072	2739	2643	3096	3113	2754	2731	2688	2835	2897	2908	3058	34534
Kerala	2003	2894	2485	2495	2570	2614	2526	2534	2569	2612	2852	2799	2997	31947
Kerala	2004	3661	3393	3477	3411	3239	3306	3164	3334	3399	3532	3425	3762	41103
Kerala	2005	3784	3435	3672	3594	3870	3410	3172	3452	3109	3467	3602	3728	42295
Kerala	2006	3901	3394	3711	3360	3600	3229	3139	3398	3384	3423	3516	3673	41728
Kerala	2007	3758	3459	3529	3481	3667	3145	2793	3107	2935	3207	3211	3569	39861
Kerala	2008	3433	3172	3121	3205	3210	2911	2904	2874	3077	3092	3034	3205	37238
Kerala	2009	3142	2933	3072	2877	3133	2967	2745	2851	2707	2910	2959	3161	35457
Kerala	2010	3221	2967	3180	3073	3149	2743	2641	2745	2861	2602	2897	3003	35082
Kerala	2011	3146	2929	3622	2805	3144	2767	2688	2787	2844	2476	2882	3126	35216
Kerala	2012	3262	3039	2960	2942	3159	2866	2848	2929	3025	2960	2991	3193	36174
Kerala	2013	3251	3023	3090	3113	3231	2712	2595	2772	2788	2728	2789	3123	35215
Kerala	2014	3332	2909	3005	2785	3162	2805	2834	2913	2911	2911	2963	3342	35872
Lakshadweep	2001	2	0	0	0	0	0	0	0	2	1	1	0	6
Lakshadweep	2002	0	0	0	0	0	0	0	0	0	0	0	0	0
Lakshadweep	2003	0	0	0	0	0	0	0	0	0	0	1	0	1
Lakshadweep	2004	0	1	0	0	0	0	0	0	0	0	1	0	2
Lakshadweep	2005	0	0	0	0	0	0	0	0	0	0	0	0	0
Lakshadweep	2006	0	0	0	0	0	0	1	0	0	0	0	0	1
Lakshadweep	2007	1	0	0	1	0	0	0	0	0	0	0	0	2
Lakshadweep	2008	0	0	0	0	0	0	0	0	0	0	0	0	0
Lakshadweep	2009	1	0	0	0	0	0	0	0	0	0	0	0	1
Lakshadweep	2010	0	0	0	0	0	0	0	0	0	0	0	0	0
Lakshadweep	2011	0	0	0	0	0	0	0	0	0	0	0	0	0
Lakshadweep	2012	0	0	0	0	0	0	0	0	0	0	0	0	0
Lakshadweep	2013	0	0	0	0	0	0	0	0	0	0	0	0	0
Lakshadweep	2014	0	0	0	0	0	0	0	0	0	1	0	0	1
Madhya Pradesh	2001	1726	1584	1764	1728	1980	1649	1428	1372	1544	1650	1711	1576	19712
Madhya Pradesh	2002	1873	1871	1921	1894	2278	1987	1776	1649	1656	1927	2088	1959	22879
Madhya Pradesh	2003	1927	1763	1891	1965	2191	2064	1779	1693	1644	1923	1886	1890	22616
Madhya Pradesh	2004	2104	2048	2002	2166	2100	1970	1868	1672	1856	2039	1882	1884	23591
Madhya Pradesh	2005	1768	1797	1830	1771	2271	1931	1601	1637	1553	1902	1749	1664	21474
Madhya Pradesh	2006	2180	1967	2144	2278	2516	2159	1929	1723	1859	2050	2211	2022	25038
Madhya Pradesh	2007	2490	2288	2409	2776	2690	2482	2114	2080	2161	2323	2384	2247	28444
Madhya Pradesh	2008	3037	2883	2917	3250	3126	2718	2699	3223	2564	2958	2449	2492	34316
Madhya Pradesh	2009	2851	2742	2723	2822	3119	2742	2449	2335	2510	2829	2549	2661	32332
Madhya Pradesh	2010	2503	2494	2435	2548	2890	2682	2428	2245	2226	2583	2635	2458	30127
Madhya Pradesh	2011	2892	2547	2722	2564	3111	2533	2370	2220	2128	2544	2526	2416	30573
Madhya Pradesh	2012	2618	2475	2413	2799	2841	2631	2146	1992	2103	2420	2522	2213	29173
Madhya Pradesh	2013	2959	2801	2953	2952	3778	3026	2535	2448	2587	2741	2621	2633	34034
Madhya Pradesh	2014	3375	3154	3470	3345	4107	3637	3104	2951	2946	3261	3288	3060	39698

Maharashtra	2001	3162	3015	3272	3332	3671	3179	2781	2755	2772	3024	3057	3249	37269
Maharashtra	2002	3424	3087	3506	3538	3970	3273	3036	2960	2881	3225	3197	3464	39561
Maharashtra	2003	3699	3373	3234	3430	3998	3584	3421	3537	3494	3774	3651	3827	43022
Maharashtra	2004	3928	3713	3677	3714	4289	3811	3430	3262	3474	3605	3693	3943	44539
Maharashtra	2005	4078	3694	3832	3877	4637	4126	3564	3410	3456	3794	3933	4185	46586
Maharashtra	2006	4062	3741	4129	4271	4881	4194	3705	3497	3705	4197	4084	4421	48887
Maharashtra	2007	4572	4318	4481	4795	5206	4400	3839	3850	3784	4090	4349	4291	51975
Maharashtra	2008	4376	4260	4479	4839	4726	4096	3738	3588	3560	3886	3854	4277	49679
Maharashtra	2009	4126	4024	4199	4373	4691	4097	3531	3737	3921	3916	3895	4355	48865
Maharashtra	2010	4383	4036	4314	4401	4861	4393	3925	3679	3938	4049	4226	4414	50619
Maharashtra	2011	4332	4052	4155	3928	4675	3976	3674	3439	3520	3511	3732	4126	47120
Maharashtra	2012	4110	3760	3896	4031	4071	3865	3539	3384	3374	3508	3523	4186	45247
Maharashtra	2013	4013	3689	3978	3798	4460	3653	3215	3081	3206	3304	3434	4032	43863
Maharashtra	2014	3978	3714	3920	3847	4258	3784	3546	3450	3298	3291	3478	3818	44382
Manipur	2001	39	31	40	48	40	14	21	26	38	41	31	41	410
Manipur	2002	51	49	44	37	41	36	50	42	42	49	39	40	520
Manipur	2003	43	40	48	43	42	44	49	48	36	44	45	41	523
Manipur	2004	34	43	41	42	40	42	20	37	36	30	45	58	468
Manipur	2005	55	57	58	48	50	45	45	38	51	47	51	55	600
Manipur	2006	50	45	51	44	44	26	36	37	35	43	61	49	521
Manipur	2007	47	46	57	51	44	42	35	31	44	44	38	59	538
Manipur	2008	56	41	39	39	34	46	32	31	45	48	42	49	502
Manipur	2009	40	54	51	41	39	39	58	37	45	58	57	59	578
Manipur	2010	74	46	49	46	41	30	44	37	54	57	62	62	602
Manipur	2011	64	66	62	63	52	58	55	54	52	50	59	56	691
Manipur	2012	67	70	64	69	57	72	62	60	58	64	58	70	771
Manipur	2013	59	46	57	58	45	46	39	40	63	67	67	84	671
Manipur	2014	61	59	66	59	62	49	49	53	73	72	59	73	735
Meghalaya	2001	11	15	13	20	12	11	15	12	9	21	12	10	161
Meghalaya	2002	31	26	43	40	39	31	28	33	43	56	51	44	465
Meghalaya	2003	33	31	31	31	34	37	33	32	26	27	29	29	373
Meghalaya	2004	22	21	29	24	31	21	27	30	32	43	28	20	328
Meghalaya	2005	38	30	38	18	27	26	23	25	25	33	17	30	330
Meghalaya	2006	16	10	17	10	14	16	10	21	15	18	13	16	176
Meghalaya	2007	19	19	13	13	12	10	13	9	18	5	15	11	157
Meghalaya	2008	13	9	21	25	24	16	11	18	12	12	15	15	191
Meghalaya	2009	28	24	36	24	37	23	19	23	18	25	19	38	314
Meghalaya	2010	23	22	31	24	20	23	20	20	20	14	19	21	257
Meghalaya	2011	24	25	33	18	20	20	30	22	12	25	13	19	261
Meghalaya	2012	33	31	28	38	30	31	29	61	22	25	14	13	355
Meghalaya	2013	20	24	20	14	20	21	16	9	11	14	13	12	194
Meghalaya	2014	30	27	20	25	27	28	29	24	26	22	19	24	301
Mizoram	2001	7	5	9	4	9	5	8	7	4	3	5	6	72
Mizoram	2002	6	0	4	9	3	2	2	4	3	0	2	6	41
Mizoram	2003	5	8	3	5	8	5	6	5	4	3	9	10	71
Mizoram	2004	6	0	4	8	3	2	2	2	3	0	2	6	38
Mizoram	2005	12	8	4	3	2	3	2	5	5	5	7	15	71
Mizoram	2006	11	4	7	6	6	5	2	6	3	3	9	10	72
Mizoram	2007	4	4	4	4	3	5	10	7	3	2	4	2	52
Mizoram	2008	6	7	8	7	6	9	8	5	5	6	12	8	87
Mizoram	2009	8	4	6	1	3	8	7	7	7	5	8	7	71
Mizoram	2010	14	18	20	8	13	9	6	6	9	6	8	8	125
Mizoram	2011	8	5	14	3	13	3	9	4	5	8	10	15	97
Mizoram	2012	7	10	7	6	8	2	20	4	12	13	6	15	110
Mizoram	2013	11	7	18	14	9	8	10	8	6	8	7	8	114
Mizoram	2014	10	6	15	10	11	4	10	5	6	3	7	8	95
Nagaland	2001	4	5	7	9	6	8	5	2	4	3	1	6	60
Nagaland	2002	5	1	1	3	3	4	2	1	3	2	2	7	34
Nagaland	2003	4	8	4	8	5	4	5	3	2	4	5	4	56
Nagaland	2004	7	8	5	7	0	3	12	6	3	7	6	4	68
Nagaland	2005	1	8	5	4	0	8	4	5	3	10	2	3	53
Nagaland	2006	7	6	8	7	6	9	6	6	5	5	4	8	77
Nagaland	2007	7	7	10	7	7	4	6	6	1	8	3	9	75
Nagaland	2008	8	10	12	8	10	13	16	3	10	10	12	14	126
Nagaland	2009	3	4	3	1	5	6	8	4	1	7	1	4	47
Nagaland	2010	10	10	2	1	2	6	1	5	2	3	1	1	44
Nagaland	2011	3	3	7	3	4	2	1	0	2	4	2	1	32
Nagaland	2012	5	4	4	4	4	4	1	1	3	3	5	4	42
Nagaland	2013	4	6	2	3	2	4	3	6	4	3	2	0	39
Nagaland	2014	3	2	2	4	2	0	1	3	0	4	0	7	28

Odisha	2001	573	546	575	593	631	528	449	460	502	486	555	508	6406
Odisha	2002	563	520	643	595	656	592	559	495	518	568	554	587	6850
Odisha	2003	558	583	594	573	619	567	566	493	500	510	530	582	6675
Odisha	2004	649	611	684	578	636	624	548	476	577	579	627	689	7278
Odisha	2005	670	684	684	679	820	594	609	544	546	538	633	592	7593
Odisha	2006	831	614	702	651	700	651	599	522	533	534	600	792	7729
Odisha	2007	835	573	722	719	720	638	671	573	569	696	719	779	8214
Odisha	2008	752	716	771	794	784	651	644	570	577	600	598	727	8184
Odisha	2009	767	756	745	740	776	846	582	675	685	702	763	855	8892
Odisha	2010	860	804	827	729	886	816	728	719	698	745	778	823	9413
Odisha	2011	841	807	890	779	932	825	763	711	547	710	736	857	9398
Odisha	2012	811	745	839	781	818	786	756	671	665	727	755	931	9285
Odisha	2013	911	806	912	765	999	809	757	677	707	619	799	919	9680
Odisha	2014	896	826	851	726	852	828	756	693	660	776	854	922	9640
Puducherry	2001	109	89	122	107	122	106	107	107	97	105	111	104	1286
Puducherry	2002	101	110	131	121	126	140	141	128	120	131	111	124	1484
Puducherry	2003	103	122	142	123	138	131	133	156	147	134	114	143	1586
Puducherry	2004	137	139	143	126	127	124	145	146	144	161	114	127	1633
Puducherry	2005	152	164	150	153	138	145	162	159	148	143	121	145	1780
Puducherry	2006	120	124	166	125	148	132	125	161	151	130	122	135	1639
Puducherry	2007	149	126	157	165	138	153	152	139	156	131	149	129	1744
Puducherry	2008	127	145	152	153	162	151	142	154	136	155	91	130	1698
Puducherry	2009	148	130	153	148	155	147	126	135	146	160	126	124	1698
Puducherry	2010	114	132	132	119	126	141	151	127	130	139	104	114	1529
Puducherry	2011	122	105	119	112	116	109	137	144	140	149	117	110	1480
Puducherry	2012	128	125	159	103	133	131	126	119	115	124	119	128	1510
Puducherry	2013	102	108	120	139	127	128	126	125	134	129	106	107	1451
Puducherry	2014	44	64	65	46	59	57	60	58	62	75	44	37	671
Punjab	2001	130	160	162	159	154	135	146	144	135	165	156	162	1808
Punjab	2002	130	128	122	115	110	130	138	123	134	118	133	127	1508
Punjab	2003	101	85	107	124	135	142	149	152	147	167	151	143	1603
Punjab	2004	167	181	163	162	160	163	166	160	195	188	170	161	2036
Punjab	2005	115	196	175	159	167	157	243	151	181	178	210	220	2152
Punjab	2006	185	182	165	163	182	200	173	187	207	211	187	209	2251
Punjab	2007	211	179	192	180	187	190	183	203	232	213	214	207	2391
Punjab	2008	230	191	176	224	197	191	190	177	184	225	221	188	2394
Punjab	2009	225	237	214	221	191	238	209	206	217	236	256	234	2684
Punjab	2010	219	195	209	219	200	200	201	187	180	196	214	219	2439
Punjab	2011	517	602	531	536	510	521	522	536	491	585	572	573	6496
Punjab	2012	525	550	529	527	502	535	505	480	483	519	610	563	6328
Punjab	2013	503	464	555	552	533	487	467	458	482	573	665	584	6323
Punjab	2014	495	493	555	524	563	535	491	519	489	568	595	564	6391
Rajasthan	2001	1733	1678	1755	1750	1937	1778	1485	1606	1562	1626	1583	1506	19999
Rajasthan	2002	1506	1567	1707	1738	1978	1878	1747	1632	1836	1716	1696	1570	20571
Rajasthan	2003	1795	1664	1707	1762	2101	1967	1837	1779	1840	1977	1907	1597	21933
Rajasthan	2004	1933	1939	1999	2124	2161	2021	1815	1697	1944	1889	2057	1664	23243
Rajasthan	2005	1804	1715	1971	1964	2429	2123	1920	1844	1843	1920	2055	1527	23115
Rajasthan	2006	1948	1926	2052	2110	2438	1998	1901	1726	1776	1984	1855	1634	23348
Rajasthan	2007	2051	1918	2054	2252	2270	1931	1822	1935	1945	2018	2084	1605	23885
Rajasthan	2008	2020	1935	1872	2111	2243	1936	1856	1829	2132	2135	2055	1580	23704
Rajasthan	2009	1983	2014	2172	2235	2438	2285	2026	2033	2084	2121	2011	1712	25114
Rajasthan	2010	2074	2024	2138	2140	2298	2163	1974	1799	1782	1961	2090	1859	24302
Rajasthan	2011	2045	1881	2006	1986	2341	2026	1859	1821	1815	1906	1957	1602	23245
Rajasthan	2012	2048	1914	1972	1978	2116	2043	1972	1773	1670	1873	1913	1697	22969
Rajasthan	2013	2034	1821	1963	2046	2305	2109	1822	1679	1944	2013	2095	1755	23586
Rajasthan	2014	2143	1951	2130	2053	2484	2213	1976	1888	1839	2035	2104	1823	24639
Sikkim	2001	9	1	6	5	3	5	0	6	6	5	7	2	55
Sikkim	2002	43	23	50	11	34	40	29	17	25	24	24	31	351
Sikkim	2003	10	10	8	11	16	16	8	13	8	7	11	13	131
Sikkim	2004	8	15	12	8	15	22	11	7	11	25	11	14	159
Sikkim	2005	10	19	13	15	17	16	8	23	12	21	20	15	189
Sikkim	2006	4	5	6	1	3	1	3	3	1	3	6	3	39
Sikkim	2007	10	7	21	7	12	12	10	12	8	16	12	23	150
Sikkim	2008	13	28	16	10	8	13	6	26	13	26	25	12	196
Sikkim	2009	57	39	28	100	43	47	29	41	46	34	50	50	564
Sikkim	2010	16	17	18	19	11	18	15	11	18	10	21	12	186
Sikkim	2011	17	9	22	19	3	7	11	8	8	12	6	7	129
Sikkim	2012	14	7	8	10	9	15	10	11	10	12	10	7	123
Sikkim	2013	16	17	33	24	24	17	18	21	17	20	16	21	244
Sikkim	2014	9	6	15	13	13	11	14	11	19	6	9	4	130

Tamil Nadu	2001	4174	4264	4724	4442	4465	4453	4370	4437	4378	3985	4261	4025	51978
Tamil Nadu	2002	4326	4300	4703	4788	4935	4609	4434	4523	4500	4273	3883	4229	53503
Tamil Nadu	2003	4278	4114	4341	3928	4244	4378	4177	4337	4555	4397	3973	4303	51025
Tamil Nadu	2004	4160	4252	4457	4421	4540	4495	4373	4696	4506	4282	4136	4190	52508
Tamil Nadu	2005	9062	8705	9205	9006	9330	5691	4235	4158	4450	4082	3601	3955	75480
Tamil Nadu	2006	4435	4322	4781	4808	5082	5029	4365	4891	4575	4320	4007	4530	55145
Tamil Nadu	2007	4855	4527	4911	5348	5414	4978	4841	5320	4850	4758	4730	4585	59117
Tamil Nadu	2008	4962	4626	5018	5465	5532	5087	4947	5436	4956	4862	4833	4685	60409
Tamil Nadu	2009	4950	4802	4927	4956	5406	5241	4975	5227	5332	5288	4723	4967	60794
Tamil Nadu	2010	5498	5199	5727	5663	6027	5393	5221	5367	5349	5426	5179	4947	64996
Tamil Nadu	2011	5522	5228	5821	5461	6018	5879	5706	5162	5688	5107	4937	5344	65873
Tamil Nadu	2012	5868	5646	5813	5819	5896	5646	5688	5815	5283	5444	5367	5472	67757
Tamil Nadu	2013	5881	5337	5741	5599	5914	5593	5311	5470	5597	5313	5268	5214	66238
Tamil Nadu	2014	5658	5293	5824	5303	5711	5915	5719	5936	5777	5516	5036	5562	67250
Tripura	2001	53	52	63	44	36	55	50	37	36	41	35	42	544
Tripura	2002	47	38	56	44	60	54	65	51	54	42	59	54	624
Tripura	2003	73	51	51	45	57	52	61	51	60	64	61	50	676
Tripura	2004	49	42	57	44	56	58	54	53	39	63	66	64	645
Tripura	2005	57	55	53	42	54	65	52	54	59	61	48	62	662
Tripura	2006	70	50	62	67	74	52	69	55	68	73	80	73	793
Tripura	2007	59	85	70	66	69	67	51	58	71	60	58	87	801
Tripura	2008	51	56	66	70	72	55	64	74	78	65	57	59	767
Tripura	2009	68	93	78	61	77	63	55	76	65	63	72	94	865
Tripura	2010	76	66	81	77	65	68	72	77	86	72	83	78	901
Tripura	2011	85	57	81	64	55	67	54	72	50	78	79	92	834
Tripura	2012	83	85	69	72	82	66	65	70	75	76	64	81	888
Tripura	2013	66	50	63	72	64	70	69	61	68	76	81	78	818
Tripura	2014	65	74	81	63	72	46	64	53	54	60	38	46	716
Uttar Pradesh	2001	1695	1737	1652	1663	1733	1027	1310	1478	1626	1439	1231	1330	17921
Uttar Pradesh	2002	1237	1164	1270	1194	1349	1363	1169	1113	1099	1100	1233	1166	14457
Uttar Pradesh	2003	1049	955	911	868	996	908	873	815	794	938	1065	1182	11354
Uttar Pradesh	2004	1309	1309	1285	1004	1174	1062	1093	1227	1164	1201	1198	1348	14374
Uttar Pradesh	2005	1266	1220	1334	1114	1422	1302	1182	1251	1080	1107	1153	1258	14689
Uttar Pradesh	2006	1608	1605	1389	1160	1280	1343	1201	1252	1226	1204	1227	1712	16207
Uttar Pradesh	2007	1647	1570	1568	1495	1511	1461	1443	1392	1394	1393	1514	1643	18031
Uttar Pradesh	2008	1799	1624	1664	1555	1478	1483	1528	1411	1489	1569	1631	1833	19064
Uttar Pradesh	2009	2010	1758	1834	1802	1912	1747	1683	1754	1694	1791	2071	1929	21985
Uttar Pradesh	2010	2213	2123	2284	2060	2171	2320	2022	1914	1862	1929	2131	2387	25416
Uttar Pradesh	2011	2045	2088	2145	1905	2178	2189	1842	1948	1879	1909	2150	2235	24513
Uttar Pradesh	2012	2101	2030	2100	2066	2380	2167	2037	1759	1670	1803	2055	2310	24478
Uttar Pradesh	2013	2290	2150	2235	2225	2554	2227	2035	1951	1871	1922	2149	2366	25975
Uttar Pradesh	2014	2550	2311	2217	2067	2265	2209	2086	2017	1824	2023	2170	2325	26064
Uttara hand	2001	78	88	83	74	94	84	86	81	82	74	84	95	1003
Uttarakhand	2002	87	76	79	106	111	121	85	97	68	94	96	97	1117
Uttarakhand	2003	80	87	78	91	120	106	113	106	84	105	106	110	1186
Uttarakhand	2004	113	99	103	97	115	122	110	81	81	105	114	112	1252
Uttarakhand	2005	120	94	96	89	129	141	106	93	112	99	124	129	1332
Uttarakhand	2006	108	118	149	124	100	116	124	115	110	135	116	146	1461
Uttarakhand	2007	133	116	124	134	148	136	119	118	131	104	116	143	1522
Uttarakhand	2008	138	126	138	119	133	125	143	73	81	108	116	109	1409
Uttarakhand	2009	99	112	101	134	115	137	124	101	91	99	133	137	1383
Uttarakhand	2010	86	108	128	127	125	145	148	128	95	129	136	124	1479
Uttarakhand	2011	114	118	119	124	135	133	112	100	115	125	149	151	1495
Uttarakhand	2012	112	152	122	137	127	138	108	123	82	118	125	111	1455
Uttarakhand	2013	84	91	118	102	167	120	94	81	104	112	97	127	1297
Uttarakhand	2014	58	43	65	64	77	62	61	72	69	83	76	71	801
West Bengal	2001	897	908	948	1039	952	972	915	990	953	823	935	1098	11430
West Bengal	2002	894	932	942	1041	911	854	915	919	922	961	940	1006	11237
West Bengal	2003	916	930	1084	975	909	980	1054	934	945	823	842	983	11375
West Bengal	2004	1044	1001	1197	1085	1086	1043	964	900	869	977	898	1088	12152
West Bengal	2005	1036	1061	1084	1035	1096	1049	998	926	887	975	979	1055	12181
West Bengal	2006	1112	1093	1187	1053	1124	1153	1011	1105	1009	1026	1081	1131	13085
West Bengal	2007	1134	1051	1275	1240	1157	1269	1176	1050	1165	1133	1068	1236	13954
West Bengal	2008	974	919	1097	1013	1069	1000	875	924	908	941	880	947	11547
West Bengal	2009	1231	1150	1183	1101	1158	1180	1162	1104	1171	1161	1176	1182	13959
West Bengal	2010	1245	1150	1349	1246	1172	1284	1231	1190	1128	1227	1251	1252	14725
West Bengal	2011	1350	1179	1314	1148	1220	1241	1185	1074	1112	1214	1161	1270	14468
West Bengal	2012	1346	1383	1357	1270	1352	1434	1349	1204	1112	1251	1179	1371	15608
West Bengal	2013	1564	1382	1474	1392	1629	1391	1315	1208	1228	1299	1335	1332	16549
West Bengal	2014	1516	1398	1473	1385	1527	1439	1416	1356	1391	1395	1415	1394	17105

Used R Code

- **For the data which has no Seasonality:**

```
library(readxl)
data=read_excel(file.choose())
head(data)
data1=data$`Uttarakhand`; data1
dat=ts(data1,start=c(2001,1),end=c(2014,12),frequency=12);dat
plot(dat)
library(tseries)
library(forecast)
acf(dat)
adf.test(dat)
X=diff(dat)
adf.test(X)
acf(X)
pacf(X)
model=ARIMA(dat,order = c(2,1,1))
model
e=model$residuals
acf(e)
Box.test(e,type = c("Ljung-Box"))
fore=forecast(model,h=36);fore
pred=data.frame(fore)
pred=pred$Point.Forecast
p=ts(pred,start = c(2015,1),end = c(2017,12),frequency = 12);p
plot(fore)
```

- **For the data which has Seasonality:**

```
library(readxl)

data=read_excel(file.choose())

head(data)

data1=data$Maharashtra;data1

dat=ts(data1,start=c(2001,1),end=c(2014,12),frequency=12);dat

plot(dat)

library(tseries)

library(forecast)

decomp=decompose(dat)

season=decomp$seasonal;season

plot(decomp)

acf(dat)

adf.test(dat)

X=data1-decomp$seasonal;X

acf(X)

adf.test(X)

d=diff(X);d

adf.test(d)

acf(d)

pacf(X)

model=ARIMA(dat,order = c(2,0,3))

model

e=model$residuals

acf(e)

Box.test(e,type = c("Ljung-Box"))

prediction=forecast(model,h=36);prediction

plot(prediction)

pred=data.frame(prediction)

pred=pred$Point.Forecast
```

```
pred

t=ts(pred,start = c(2015,1),end = c(2017,12),frequency = 12);t

vec=c(56.157853,-3.005609,101.109776,18.487981,47.885417,38.244391,-18.079327,-69.479968,-81.887019,-53.755609,-63.322917,27.645032)

p=vec+pred

r=ts(p,start = c(2015,1),end = c(2017,12),frequency = 12);r
```

References

1. **C. Chatfield:** The Analysis of Time Series – An Introduction.
2. **G. E. P. Box and G. M. Jenkins:** Time Series Analysis–Forecasting and Control.
3. **P. J. Brockwell and R. A. Davis:** Introduction to Time Series and Forecasting.
4. **A. Pankratz:** Forecasting with Univariate Box-Jenkins Model.
5. **D. C. Montgomery, C. L. Jennings and M. Kulahci:** Introduction to Time Series Analysis and Forecasting