

Abhitejh Das (MSc Data Science (DS))

# Time Series Analysis on Crime and Drug Use

Presented to LJMU University

Research Proposal submission

Topic Time Series Analysis for Crime and Drug Use

By

Abhitejh Das

Student ID: **PN1030500**

April 2023

## **Abstract**

India has struggled with crime for many years, but things have changed significantly. However, crime continues to be one of the major social issues. We must make use of technology from the 21st century to create safer society. It is possible to analyse crime patterns and predict future criminal activity thanks to current technologies and data availability. Police can utilise this data to tighten security and notify the neighborhood's residents. One such component being used in India by the Police Authorities is "predictive policing." This study examines, contrasts, and projects future crime rates based on historical patterns of crime in the States of India. If individuals are seeking for something specific, these results may aid them and provide them options.

## **Introduction**

According to Time Series Analysis and Forecasting, any data that is periodically recorded over time can be utilised to predict linked future occurrence.

This project examines crime statistics in Victims of Rape and provides numerous visualisations for simple comprehension of the findings.

## **Technologies Used**

**Python:** Python may be tall level Programming language It may be general-purpose programming tongue with affluent library back for machine learning models and stat models. Evaluating models inspected in this amplify are imported from 'stats-models ' python library. Another basic library for the wander is 'pandas', which has diverse strategies to work with gigantic data.

**Pandas :** Python Pandas is one of the foremost dependable library when it comes to taking care of huge data sets. Its performance and instinct has made it one of the foremost well known libraries accessible for data investigation. There could be other libraries out there but 'pandas' is exceptionally simple to utilize and work with.

## **R programming Dialect**

R is an open source dialect which gives bigger bolster for measurable investigation, and specialization in it, whereas Python gives an object-oriented approach and stunning number of integrative with other modules.

## **Stats models:**

Python libraries incorporate 'stats models' which gives capacity for the assessment and estimation of distinctive factual models, for performing factual tests and data investigation. 'Auto Backward Moving

Window' ("ARIMA"), Auto "ARIMA", Holt's Winter Exponential smoothing models examined in this project from stats show library.

**Anaconda:** Boa constrictor is free and simple to introduce bundle director for python. It created conditions to run python records with different machine learning libraries. Because it can keep up all the specified libraries, bundles for software engineer, it is much less difficult for software engineer to preserve the improvement conditions.

#### **Jupyter Scratch pad:**

Jupyter note pad is open source webapplication that allows programmer to preserve code, description, comments, and visualizations at single put. It is exceptionally valuable for Machine Learning ventures as engineers can see the visualizations and code at the same put. It is client inviting and simple to begin with.

#### **Dataset :**

Data set utilized contains total data approximately different perspectives of wrongdoings happened in India from 2001 totally different states. There are numerous variables that can be analyzed from this dataset. Over all, I trust this dataset makes difference us to get it way better around India.

#### **Analysis of Data:**

Data Processing

```
Crime_2018 = read_csv("../input/crime-in-india/Victims of Rape/2018 Victims of Rape.csv")
```

#### — Column specification —

```
cols(
  `S. No` = col_double(),
  Category = col_character(),
  `State/UT` = col_character(),
  `Cases Reported` = col_double(),
  `Child Victims of Rape (Below 18 Yrs) - Below 6 Years` = col_double(),
  `Child Victims of Rape (Below 18 Yrs) - 6 Years & Above` = col_double(),
  `Child Victims of Rape (Below 18 Yrs) - 12 Years & Above` = col_double(),
  `Child Victims of Rape (Below 18 Yrs) - 16 Years & Above` = col_double(),
  `Child Victims of Rape (Below 18 Yrs) - Total Girl/Child Victims` = col_double(),
  `Women Victims of Rape (Above 18 Yrs) - 18 Years & Above - Below 30 Years` = col_double(),
  `Women Victims of Rape (Above 18 Yrs) - 30 Years & Above - Below 45 Years` = col_double(),
  `Women Victims of Rape (Above 18 Yrs) - 45 Years & Above - Below 60 Years` = col_double(),
  `Women Victims of Rape (Above 18 Yrs) - 60 Years & Above` = col_double(),
  `Women Victims of Rape (Above 18 Yrs) - Total Women/Adult Victims` = col_double(),
  `Total Victims` = col_double()
)
```

Activ

```
glimpse(crimeindia)
```

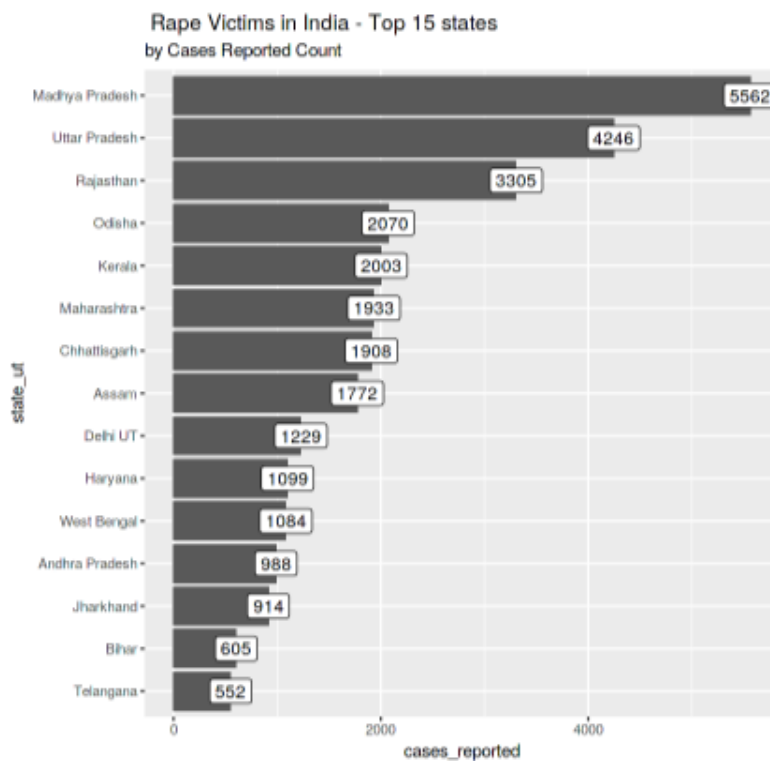
```
Rows: 36
Columns: 15
$ `S. No`                                <dbl> ...
$ Category                               <chr> ...
$ `State/UT`                             <chr> ...
$ `Cases Reported`                       <dbl> ...
$ `Child Victims of Rape (Below 18 Yrs) - Below 6 Years` <dbl> ...
$ `Child Victims of Rape (Below 18 Yrs) - 6 Years & Above - Below 12 Years` <dbl> ...
$ `Child Victims of Rape (Below 18 Yrs) - 12 Years & Above - Below 16 Years` <dbl> ...
$ `Child Victims of Rape (Below 18 Yrs) - 16 Years & Above - Below 18 Years` <dbl> ...
$ `Child Victims of Rape (Below 18 Yrs) - Total Girl /Child Victims` <dbl> ...
$ `Women Victims of Rape (Above 18 Yrs) - 18 Years & Above - Below 30 Years` <dbl> ...
$ `Women Victims of Rape (Above 18 Yrs) - 30 Years & Above - Below 45 Years` <dbl> ...
$ `Women Victims of Rape (Above 18 Yrs) - 45 Years & Above - Below 60 Years` <dbl> ...
$ `Women Victims of Rape (Above 18 Yrs) - 60 Years & Above` <dbl> ...
$ `Women Victims of Rape (Above 18 Yrs) - Total Women/Adult Victims` <dbl> ...
$ `Total Victims`                       <dbl> ...
```

```
head(crimeindia)
```

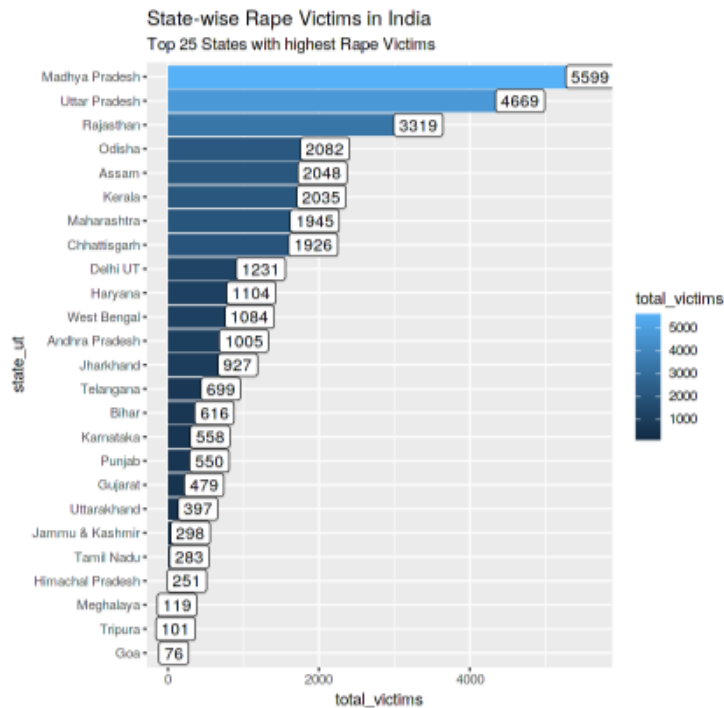
A tibble: 6 × 15

s_no	category	state_ut	cases_reported	child_victims_of_rape_below_18_yrs_below_6_years	child_victims_of_rape_below_18_yrs_6_years_above_below_
<dbl>	<chr>	<chr>	<dbl>		<dbl>
1	State	Andhra Pradesh	988		14
2	State	Arunachal Pradesh	59		1
3	State	Assam	1772		2
4	State	Bihar	605		0
5	State	Chhattisgarh	1908		47
6	State	Goa	76		4

Raw data before pre-processing



Crime count per each state



Victim Count per each state

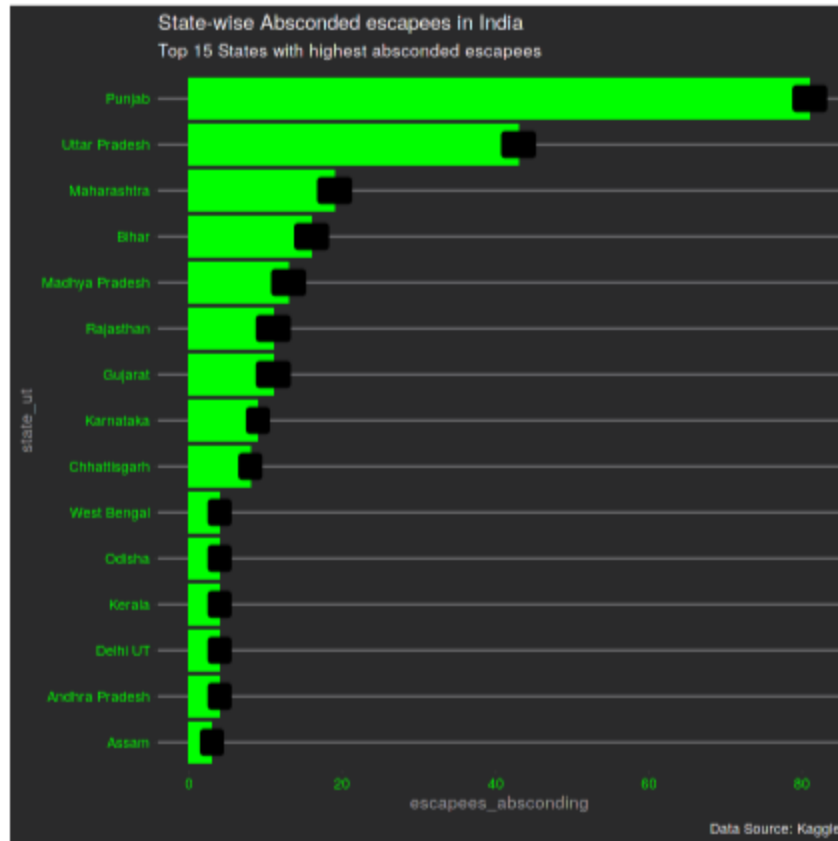
the top 15 states with highest absconded escapees in India

```

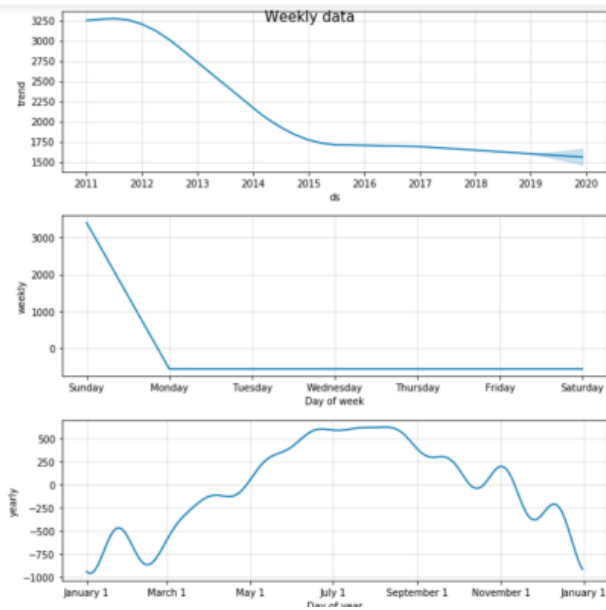
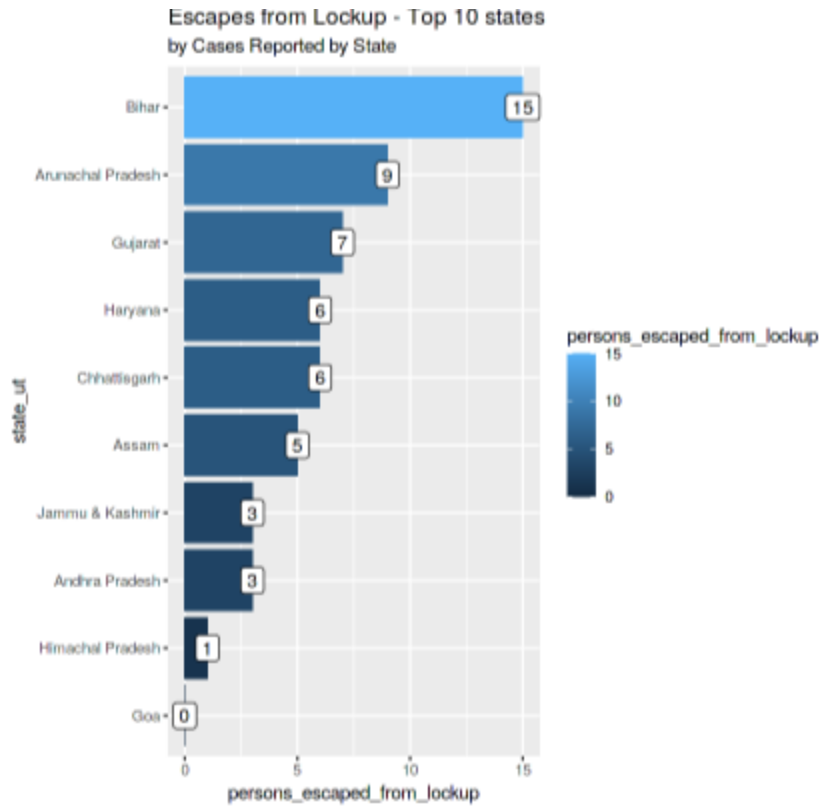
policecustodyescape %>%
  select(one_of('state_ut', 'escapees_absconding')) %>%
  arrange(desc(escapees_absconding)) %>%
  head(15) %>%
  mutate(state_ut = fct_reorder(state_ut, escapees_absconding)) %>%
  ggplot() + geom_col(aes(y = state_ut, x = escapees_absconding), fill = 'green') +
  geom_label(aes(y = state_ut, x = escapees_absconding, label = escapees_absconding), fill = 'white')

labs(title = 'State-wise Absconded escapees in India',
      subtitle = 'Top 15 States with highest absconded escapees',
      caption = 'Data Source: Kaggle') +
ggthemes::theme_hc(
  base_size = 10,
  base_family = "sans",
  style = c("darkunica")
) +
theme(axis.text.x=element_text(colour="green"),
      axis.text.y=element_text(colour="green"))

```

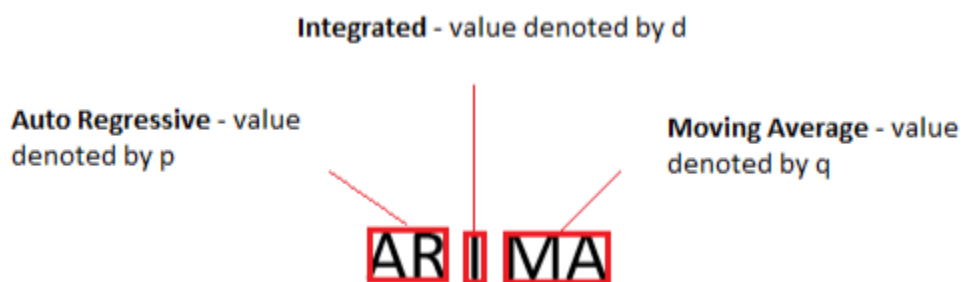


States with highest absconded escapees in India





## Building and Comparing “Arima” Model, FB Pro-phet Method :



```
# print predictions  
prediction
```

7]:

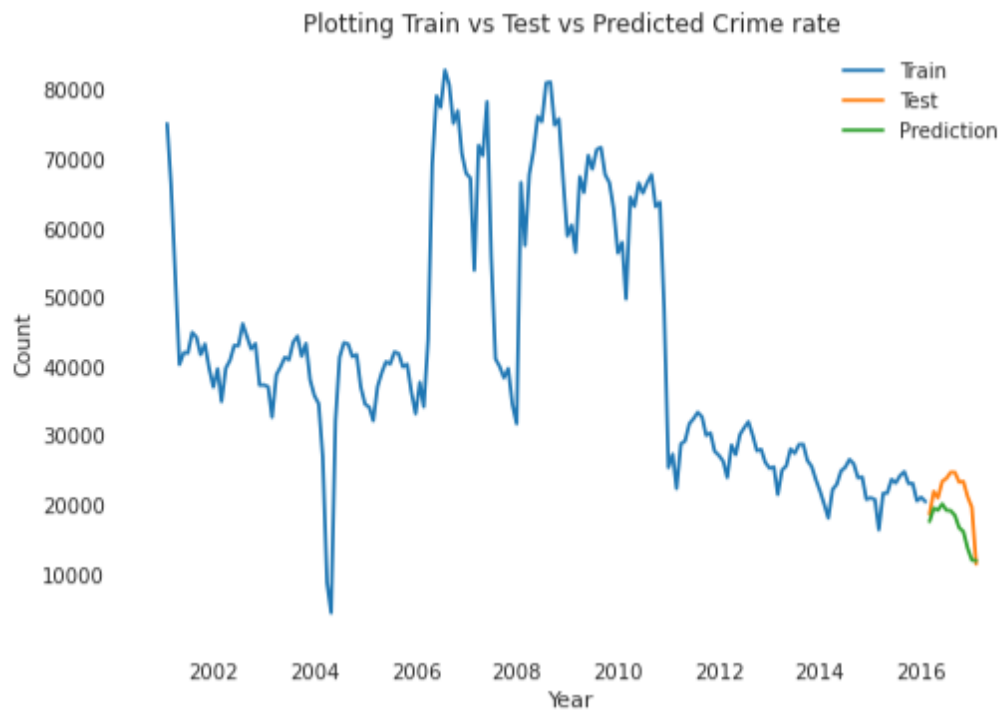
Predicted Crime Count	
Date	
2016-02-29	17475.090836
2016-03-31	19389.553070
2016-04-30	19110.124824
2016-05-31	20006.479958
2016-06-30	19100.002586
2016-07-31	19022.283472
2016-08-31	18331.380821
2016-09-30	16567.773829
2016-10-31	16002.337614
2016-11-30	13621.423353
2016-12-31	11894.500183
2017-01-31	11802.136150

“Arima” Model Summary

48]:

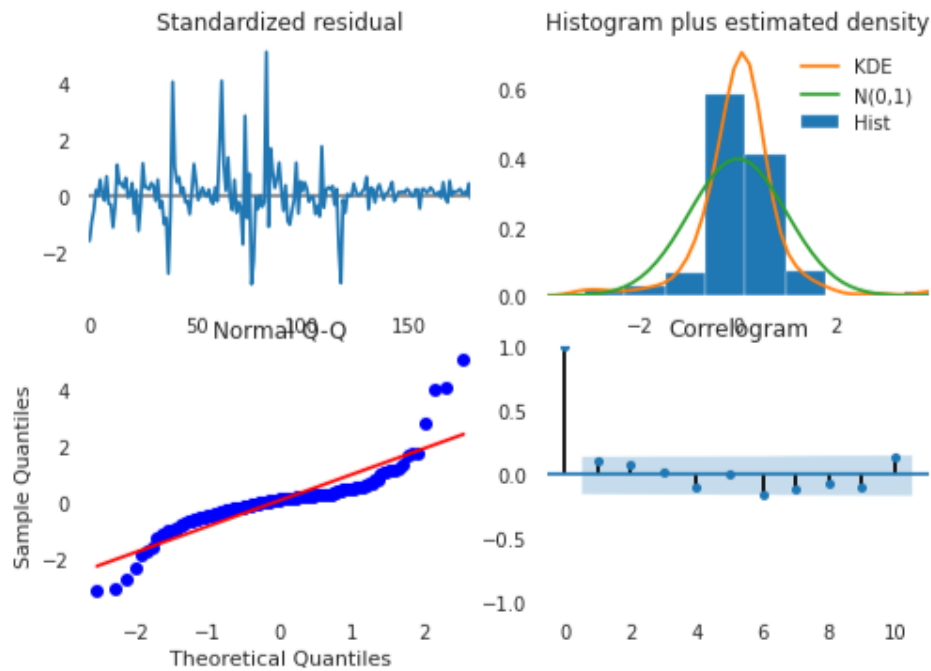
	Crime Count	fbprophet_error
Date		
2016-02-29	18537	8932.919290
2016-03-31	21814	9008.116122
2016-04-30	20885	8176.223573
2016-05-31	23269	5125.111196
2016-06-30	23701	6304.798363
2016-07-31	24577	6192.392132
2016-08-31	24550	7471.842071
2016-09-30	23176	9816.247661
2016-10-31	23243	9880.184348
2016-11-30	21077	12674.197823
2016-12-31	19517	15206.797135
2017-01-31	11329	1820.084051

Facebook Prophet Model



Plotting Train vs Test vs Predicted Crime rate

```
out = model.plot_diagnostics()
```



Time arrangement investigation may be particular way of analyzing grouping of data focuses collected over an interim of time. In time arrangement examination, examiners record data focuses at steady intervals over set period of time instead of fair recording the data focuses irregularly or haphazardly.

Time Series Analysis offer assistance in understanding the fundamental patterns, regularity and designs within the data. As the variable is time subordinate, patterns and regularity alter at the side time. Estimating of future occasions can be performed on such data which is subordinate on Time.

### **Time Series Analysis Method: "ARIMA":**

"ARIMA" may be evaluating strategy that gages long run values of time course of action based completely on its have inertia. "ARIMA" is common illustrate which is exact adequate to remove remaining autocorrelation. The input time course of action to "ARIMA" illustrate got to be stationary time course of action and this stationary is fulfilled by either differencing or logging. time course of action is said to be stationary within the occasion that its pitiless (typical), standard deviation, variance, auto relationship etc. are standard with time.

The essential step for actualizing "ARIMA" is to form the course of action stationary. It is basic to find that given course of action is stationary or non-stationary. As indicated, stationary course of action have reliable unfeeling, variance with time. They are reasonable subjective course of action like white commotion. On the other hand, non-stationary course of action have float and consistency

.

The preeminent basic step in "ARIMA" is choosing the orchestrate of the "ARIMA" appear. In common, it is said that p, d, q values shows the organize of "ARIMA". 'p' illustrates AR (Auto In reverse) component, it portrays number of past values utilized to figure long term regard. 'd' is the level of logging or differencing inside the component. This degree of differencing makes the course of action stationary. 'q' states the goof inside the illustrate as aggregation of past botch values. Auto in reverse, differencing and moving ordinary make up non-seasonal "ARIMA" appear as coordinate condition.

**FB Prophet:** This method is determining show for time arrangement data which can handle annually, week after week, and every day regularity counting occasion impacts. When the data is time subordinate and has tall history for regularity prophet is the leading show to figure. Concurring to Prophet documentation in GitHub, Facebook employments prophet for numerous dependable figures strong to exceptions and lost data. Prophet API is accessible both is R and Python for determining.

### **Holt's Winter Forecasting:**

This procedure to boot known as 'Triple Exponential Smoothing'. Essential Exponential Smoothing, twofold exponential smoothing can be utilized to figure time course of action. Triple Exponential Smoothing is more suitable for data with tall normality and incline. Planning Estimations handbook says that past recognitions are weighted so also in single moving midpoints. In separate in exponential higher regard weights are allotted to afterward values.

## **Outcome/Conclusion**

Time Arrangement Examination and Estimating is performed with few visualizations and factual models. Agreeing to estimating Madhya Pradesh and Express Pradesh have most Assault wrongdoings and Punjab Sate has tall Departing suddenly cases.

By

**Abhitejh Das**, Senior Software Engineer at Accenture Solutions PVT Limited, PG Diploma in Business Analytics, Masters Student at LJM University.