

## **Data cleaning and Visualization using R**

**# Checking table and format.**

```
> str(indian_crime_dataset)
```

**# Changing the format of date\_reported, date\_occured, date\_case\_closed from character to date time.**

```
indian_crime_dataset$date_reported <-  
as.POSIXct(indian_crime_dataset$date_reported, format = "%d-%m-%Y %H:%M")
```

```
indian_crime_dataset$date_of_occurrence <-  
as.POSIXct(indian_crime_dataset$date_of_occurrence, format = "%d-%m-%Y  
%H:%M")
```

```
indian_crime_dataset$date_case_closed <-  
as.POSIXct(indian_crime_dataset$date_case_closed, format = "%d-%m-%Y  
%H:%M")
```

**# Changing the format of time\_of\_occurrence from character to time.**

```
indian_crime_dataset$time_of_occurrence <-  
format(as.POSIXct(indian_crime_dataset$time_of_occurrence, format =  
"%d-%m-%Y %H:%M"), "%H:%M:%S")
```

**#Creating a new column named year\_of\_occurrence using, date\_of\_occurrence column.**

```
indian_crime_dataset$year_of_occurrence <-  
format(indian_crime_dataset$date_of_occurrence, "%Y")
```

**# Removing rows with null date\_of\_occurrence**

```
indian_crime_dataset <-  
indian_crime_dataset[!is.na(indian_crime_dataset$date_of_occurrence),]
```

**# Final check for data correction**

```
str(indian_crime_dataset)  
summary(indian_crime_dataset)
```

## **Starting Visualization**

### **# Installing packages**

```
install.packages("ggplot2")
library(ggplot2)

install.packages("tidyverse")
library(tidyverse)

install.packages("dplyr")
library(dplyr)
```

### **# Total Crimes as per City**

```
ggplot(indian_crime_dataset, aes(x = city)) +
  geom_bar(fill = "#2ca02c") +
  theme_minimal() +
  labs(title = "Crime Count by City")
```

### **# Crime trend over year**

```
crime_year <- indian_crime_dataset %>%
  group_by(year_of_occurrence) %>%
  summarise(total_crime = n())

ggplot(crime_year, aes(x = year_of_occurrence, y = total_crime)) +
  geom_line(color = "#d62728", linewidth = 1) +
  geom_point(color = "#d62728") +
  theme_minimal() +
  labs(title = "Crime Trend Over Years")
```

### **# Crime Type Distribution**

```
ggplot(indian_crime_dataset, aes(x = crime_description)) + geom_bar() +
  coord_flip() + theme_minimal() + labs(title = "Crime type Distribution")
```

### **# Crime Type Distribution**

```
ggplot(indian_crime_dataset, aes(x = crime_description)) +
  geom_bar(fill = "#9467bd") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Crime Type Distribution")
```

### **# Crime Domain by City**

```
ggplot(indian_crime_dataset, aes(x = city, fill = crime_domain)) +  
  geom_bar() +  
  scale_fill_brewer(palette = "Set2") +  
  theme_minimal() +  
  labs(title = "Crime Domain by City")
```

### **# Victim Gender Distribution**

```
ggplot(indian_crime_dataset, aes(x = victim_gender, fill = victim_gender))  
+  geom_bar() +  
  scale_fill_manual(values = c("M" = "#1f77b4", "F" = "#ff7f0e")) +  
  theme_minimal() +  
  labs(title = "Victim Gender Distribution")
```

### **# Weapon Used Frequency**

```
ggplot(indian_crime_dataset, aes(x = weapon_used)) +  
  geom_bar(fill = "#8c564b") +  
  coord_flip() +  
  theme_minimal() +  
  labs(title = "Weapons Used in Crimes")
```

### **# Crime by Hour of Day**

```
ggplot(indian_crime_dataset, aes(x = time_of_occurrence)) +  
  geom_bar(fill = "#17becf") +  
  theme_minimal() +  
  labs(title = "Crime Occurrence by Hour")
```

### **# Victim Age by Crime Type**

```
ggplot(indian_crime_dataset, aes(x = crime_description, y = victim_age,  
fill = crime_description)) +  
  geom_boxplot() +  
  coord_flip() +  
  theme_minimal() +  
  theme(legend.position = "none")
```

```
# City vs Year
crime_heat <- indian_crime_dataset %>%
  group_by(city, year_of_occurrence) %>%
  summarise(count = n(), .groups = "drop")

ggplot(crime_heat, aes(x = year_of_occurrence, y = city, fill = count)) +
  geom_tile() +
  scale_fill_gradient(low = "lightblue", high = "darkred") +
  theme_minimal() +
  labs(title = "Crime Heatmap")
```